

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



Máster Universitario en Bioinformática y Biología
Computacional

TRABAJO FIN DE MÁSTER

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

Autor: BARRERO RODRÍGUEZ, Rafael

Cotutora: FERRARINI, Alessia

Cotutora: MASTRANGELO, Annalaura

Ponente: LÓPEZ CORCUERA, Beatriz

Enero de 2021

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

Autor: BARRERO RODRÍGUEZ, Rafael

Cotutora: FERRARINI, Alessia
Laboratorio de Proteómica Cardiovascular

Cotutora: MASTRANGELO, Annalaura
Laboratorio de Inmunobiología

Ponente: LÓPEZ CORCUERA, Beatriz

Centro Nacional de Investigaciones Cardiovasculares (CNIC)

Enero de 2021

Resumen

La metabolómica es el estudio sistemático del perfil metabólico de las muestras biológicas. En concreto, el presente trabajo se ha centrado en los estudios de metabolómica no dirigida que utilizan como método analítico la cromatografía de líquidos acoplada a la espectrometría de masas con analizadores de alta resolución. La última etapa en el procesado de los datos obtenidos mediante esta estrategia analítica consiste en la identificación, por diferentes niveles de confianza, de los metabolitos de interés. Para ello, el investigador procede, en primera instancia, con la anotación de los analitos detectados mediante la búsqueda de sus masas exactas en bases de datos públicas. Por cada masa buscada, el investigador obtiene una lista con múltiples identificaciones, ninguna de las cuales permite identificar el analito de interés con un alto nivel de confianza. Sin embargo, aunque las anotaciones no permitan realizar una identificación inequívoca, el investigador puede aumentar el nivel de confianza y fiabilidad de la identificación mediante la revisión manual de cada una de las posibles anotaciones. Este proceso permite interpretar los resultados del estudio (si no se realizan análisis adicionales) o priorizar la compra de estándares químicos y de los análisis adicionales mediante espectrometría de masas en tándem (si se precisa un mayor nivel de confianza en las identificaciones). Sin embargo, dado el elevado número de anotaciones y la ausencia de herramientas que automaticen el proceso, este paso es lento y laborioso para el investigador.

En este proyecto hemos desarrollado un conjunto de programas que agilizan el manejo y la simplificación de las matrices con las posibles identificaciones a efectos de facilitar su visualización, la extracción de información útil y, en última instancia, la priorización de los metabolitos candidatos. Asimismo, dichos programas fueron integrados bajo una aplicación de escritorio y una aplicación web, ambas de código abierto, que permiten al usuario diseñar flujos de trabajo personalizados de manera sencilla e intuitiva.

Palabras Clave

Anotación de metabolitos, bioinformática, identificación de metabolitos, LC-MS, metabolómica.

Abstract

Metabolomics is the systematic study of the metabolic profile of biological samples. Specifically, the presented work is based on data obtained by untargeted metabolomics studies that use liquid chromatography coupled to mass spectrometry with high-resolution analysers as the analytical platform for the data acquisition. In this context, the final step of the data processing consists of the identification, by different levels of confidence, of the metabolites of interest. Firstly, the researcher annotates the detected analytes by searching for their accurate mass against public databases. For each mass searched in this step, the researcher obtains a list of multiple identifications, none of which possibly leading to a high confident identification of the analyte of interest. However, although annotation does not provide the unambiguous identification, the researcher can increase the confidence and reliability of the identification by manually reviewing each possible annotation. This step enables the researcher either interpreting the results of the study (if no additional analysis is carried out) or prioritizing the purchase of chemical standards and the additional analyses in tandem MS (if a higher level of confidence of the identification is requested). However, given the high number of annotations and the lack of tools to automate the process, this step entails a long and demanding task for the researcher.

In this project, we have developed a set of programs that streamline the data matrix handling and simplification, facilitating data visualization, useful information mining and, eventually, the prioritization of which candidate metabolites to investigate further. Likewise, these programs were integrated under a desktop application and a web application, both open source, which allow the user to design customized workflows in a simple and intuitive way.

Key words

Bioinformatics, LC-MS, metabolite annotation, metabolite identification, metabolomics.

Agradecimientos

Al Centro Nacional de Investigaciones Cardiovasculares (CNIC) y a ACCIONA, por su Programa de Becas Máster en Biomedicina, que ha sido un apoyo económico indispensable para el estudio del Máster y la realización de este trabajo.

A las codirectoras de este proyecto, Alessia y Annalaura, por su ayuda, paciencia y dedicación en este trabajo, y por ser un ejemplo de profesionalidad y rigor.

A José Manuel Rodríguez Carrasco, por todos sus consejos y recomendaciones, que han servido de guía esencial en el desarrollo de este proyecto.

A Jesús Vázquez y a todos los miembros del Laboratorio de Proteómica Cardiovascular y de la Unidad de Proteómica y Metabolómica del CNIC, por el apoyo y los ánimos recibidos.

A todos los profesores del Máster en Bioinformática y Biología Computacional de la Universidad Autónoma de Madrid, en especial a Eduardo Pérez Pérez.

A Lucía, a Julia y a mis padres, por su lucha incansable durante estos meses tan difíciles.

Finalmente, a todos los trabajadores esenciales, que durante la pandemia de COVID-19 hicieron todo lo posible para mitigar los efectos de la crisis sanitaria.

Índice general

Índice de Figuras	IX
Índice de Tablas.....	XI
1. Introducción.....	1
1.1. Marco Científico del Proyecto: La Metabolómica	1
1.2. Flujo de Trabajo: Procesamiento de datos obtenidos con LC/ESI-MS	3
1.3. Motivación del Proyecto: Límites en la identificación de los metabolitos.....	6
2. Diseño y Desarrollo	9
2.1. Introducción: Estructura del programa	9
2.1.1. Tagger: Clasificación de metabolitos	9
2.1.2. RENAME: Eliminación de información redundante.....	13
2.1.3. RowMerger: Combinación de anotaciones.....	17
2.1.4. TableMerger: Combinación de matrices	18
2.2. Integración de los módulos: Aplicación de escritorio	19
2.3. Integración de los módulos: Aplicación web.....	24
3. Pruebas Realizadas y Resultados Obtenidos	29
3.1. Introducción: Obtención de la matriz con las anotaciones putativas.....	29
3.2. Simplificación automática de la matriz usando TurboPutative	30
4. Conclusiones y Trabajo Futuro	35
4.1. Conclusiones.....	35
4.2. Trabajo Futuro	36
Glosario de acrónimos y abreviaturas	37
Bibliografía.....	39
Apéndice A. Tablas Procesadas con las Anotaciones Putativas.....	43
Apéndice B. Manual de Instalación.....	97
Apéndice C. Organización del Código Fuente	99

Índice de Figuras

1.2.1. Flujo de trabajo de los experimentos de metabolómica no dirigida	3
1.2.2. Representación 3D de todas las señales detectadas mediante LC/MS	4
2.1.1.1. Ejemplo de la clasificación de metabolitos contenidos en una tabla empleando el módulo <i>Tagger</i>	13
2.1.2.1 Procesado del nombre de metabolitos utilizando el sistema de expresiones regulares empleado por el módulo <i>REname</i>	14
2.1.2.2. Mecánica del procesamiento de los nombres de los metabolitos con las expresiones regulares.....	15
2.1.2.3. Combinación de péptidos con la misma composición aminoacídica usando el módulo <i>REname</i>	16
2.1.2.4. Procesado del nombre de lípidos isómeros con ácidos grasos usando el módulo <i>REname</i> y el paquete Goslin.....	17
2.1.3.1. Combinación de anotaciones usando el módulo <i>RowMerger</i>	18
2.1.4.1. Combinación de la información de las <i>features</i> usando <i>TableMerger</i>	19
2.2.1. Interfaz de la aplicación de escritorio de TurboPutative	21
2.2.2. Diseño y ejecución del flujo de trabajo utilizando aplicación de escritorio de TurboPutative	23
2.3.2. Sección <i>Execute</i> del sitio web	25
2.3.3. Sección de espera y descarga del sitio web	26
2.3.4. Diagrama de las consultas realizadas en la página de espera del sitio web.....	27
3.2.1. Reducción del número de filas de las tablas con las anotaciones putativas usando TurboPutative	31

Índice de Tablas

2.1.1.1. Expresión regular utilizada por el módulo <i>Tagger</i> para identificar los compuestos con elementos halógenos	11
2.1.1.2. Expresión regular utilizada por el módulo <i>Tagger</i> para identificar los compuestos peptídicos.....	12
3.1.1. Parámetros utilizados para realizar la búsqueda de posibles identificaciones en CMM	30
A.1. Extracto de la tabla generada por el módulo <i>Tagger</i>	44
A.2. Extracto de la tabla generada por el módulo <i>REname</i>	76
A.3. Extracto de la tabla generada por el módulo <i>RowMerger</i>	82
A.4. Extracto de la tabla generada por el módulo <i>TableMerger</i>	86
A.5. Priorización y descarte de las posibles identificaciones asignadas a las distintas features	91

1

Introducción

1.1. Marco científico del Proyecto: La Metabolómica

Los avances en tecnología científica y poder computacional producidos en las últimas décadas han favorecido el desarrollo de disciplinas conocidas como ciencias *ómicas*. Este campo abarca áreas como la genómica, la proteómica y la metabolómica, que se caracterizan por aplicar un enfoque holístico en el estudio de los sistemas biológicos, mediante la realización de experimentos de alto rendimiento en los que se genera una ingente cantidad de información. La necesidad de procesar y analizar los datos generados en estos estudios ha dado pie al desarrollo de la bioinformática, una disciplina que pretende aplicar el conocimiento existente en diversos campos, como la informática, la estadística o la matemática, para la resolución de los problemas biológicos.

De todas las disciplinas *ómicas*, el presente proyecto se desarrolla en el marco de la metabolómica, la disciplina *ómica* que emplea avanzadas técnicas de química analítica para caracterizar el perfil metabólico de las muestras biológicas. Del mismo modo que la transcriptómica y la proteómica estudian la totalidad de los transcritos o de las proteínas presentes en una muestra, el objeto de estudio de la metabolómica son los metabolitos, es decir, los compuestos químicos de bajo peso molecular (hasta 1,500 Da) que participan en las reacciones bioquímicas que tienen lugar en los sistemas biológicos [1]. Con frecuencia se ha descrito la metabolómica como la disciplina más apropiada para la descripción del fenotipo molecular de un sistema biológico [2]. De hecho, los estudios en este campo permiten analizar el perfil metabólico de las muestras biológicas y sus alteraciones ante un determinado trastorno o tratamiento. Estos estudios pueden tener aplicaciones muy relevantes en diferentes áreas como la bioquímica, la fisiología o la biomedicina.

Los experimentos de metabolómica pueden tener dos enfoques: dirigido y no dirigido. Los análisis dirigidos centran su estudio en un reducido número de metabolitos

seleccionados *a priori*, mientras que los análisis no dirigidos tienen como objetivo medir la totalidad de los metabolitos presentes en una muestra sin hipótesis previa.

Las dos plataformas analíticas más utilizadas en metabolómica son la resonancia magnética nuclear (del inglés *nuclear magnetic resonance*, NMR) y la espectrometría de masas (del inglés *mass spectrometry*, MS). Muchos de los estudios no dirigidos utilizan como método analítico la MS acoplada a técnicas previas de separación como la cromatografía de gases (GC/MS), la cromatografía de líquidos (LC/MS) o la electroforesis capilar (CE/MS). La técnica de ionización más usada en GC/MS es la ionización electrónica (del inglés *electron ionization*, EI), que causa la fragmentación extensa de la molécula caracterizada por espectros de fragmentación altamente reproducibles, los que permiten, a través de la comparación de dichos espectros con aquellos incluidos en las bases de datos (p. ej. NIST y Wiley), la identificación de los metabolitos detectados. Por otro lado, en el caso de LC/MS y CE/MS, la técnica de ionización más utilizada es la ionización por electrospray (del inglés *electrospray ionization*, ESI), que produce una ionización suave, lo que evita la fragmentación del compuesto y deja intacto el ion molecular. En estos casos, la identificación de los metabolitos se realiza por distintas etapas, lo que incluye, en primera instancia, la comparación de la masa monoisotópica de los analitos detectados con la masa de los compuestos incluidos en las bases de datos (p. ej. HMDB, Metlin, Kegg, entre otros) [3]–[6]. Por ello, sobre todo en la metodología no dirigida, donde no se conocen los metabolitos a analizar *a priori*, y en el caso de emplear LC/MS y/o CE/MS para el análisis, se requieren espectrómetros de masas con analizadores de alta resolución como el TOF (del inglés *time of flight*) o el Orbitrap, que permiten detectar la masa monoisotópica del analito con una precisión de hasta 0.0001 Da. Dichos analizadores permiten reducir el número de posibles anotaciones a partir de una misma masa monoisotópica. Asimismo, para reducir el número de posibles anotaciones, es frecuente el uso de equipos híbridos que acoplan analizadores de alta resolución, como el Orbitrap o el TOF, a analizadores de menor resolución, como la trampa de iones o el cuadrupolo. Esto permite la posible identificación de los metabolitos tras su fragmentación mediante el empleo de una energía de colisión y la comparación de su patrón de fragmentación con el de los estándares sintéticos analizados bajo las mismas condiciones [7], [8].

El presente proyecto se ha centrado en los estudios de metabolómica no dirigida que utilizan como método analítico la LC/MS con analizadores de alta resolución

(Orbitrap) y ESI como técnica de ionización. Como explicábamos anteriormente, esta estrategia analítica se caracteriza por una elevada complejidad en la etapa de identificación de los metabolitos, y, por lo tanto, será el foco del presente proyecto.

1.2. Flujo de Trabajo: Procesamiento de datos obtenidos con LC/ESI-MS

El flujo de trabajo de los experimentos de metabolómica no dirigida consta de una serie de pasos que abarcan desde la recogida de la muestra hasta la interpretación de los resultados (Figura 1.2.1) [9]. Por su relevancia en este proyecto, nos centraremos en la etapa de procesamiento de los datos y concretamente en la etapa de identificación de los metabolitos.

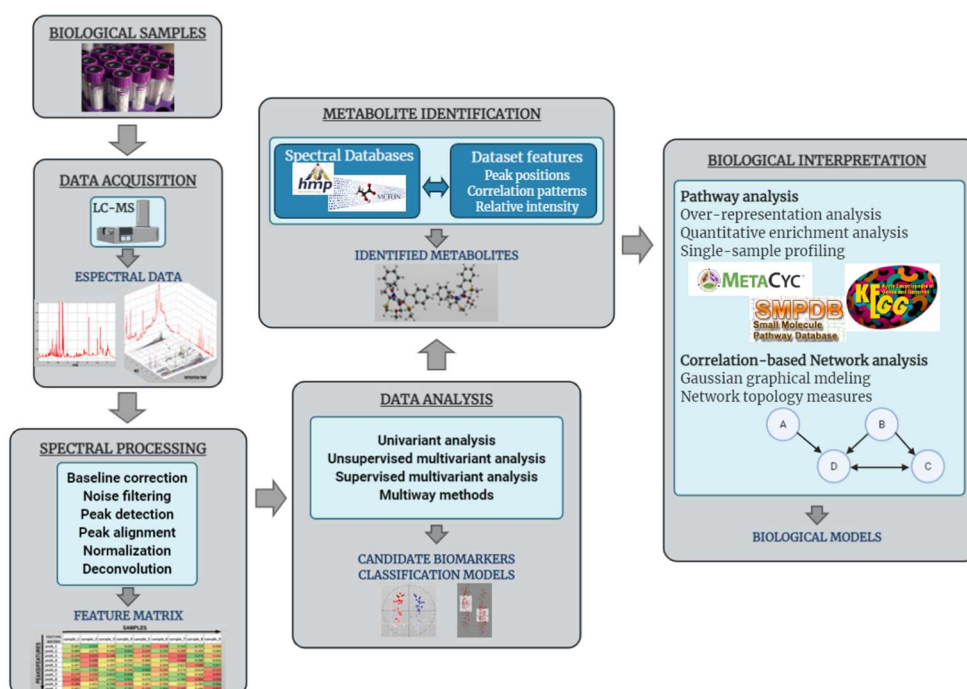


Figura 1.2.1. Flujo de trabajo de los experimentos de metabolómica no dirigida. En primer lugar, se realiza la recogida de la muestra y su pretratamiento, para posteriormente analizarlas mediante el método analítico empleado. Los datos espectrales obtenidos son procesados para realizar los análisis estadísticos y la identificación de los metabolitos. Finalmente, se procede con la interpretación biológica de los resultados. Figura adaptada de [9].

Los datos primarios obtenidos en los estudios basados en LC/MS tienen una estructura tridimensional (*i.e.* tiempo de retención *vs* relación m/z *vs* intensidad), de tal forma que los espectros generados en los distintos tiempos de retención (RT) contienen diferentes relaciones m/z con una intensidad asociada (Figura 1.2.2) [10]. Gran parte de las señales detectadas son originadas por metabolitos, artefactos, contaminantes y ruido

de fondo. Asimismo, cada metabolito puede generar múltiples señales debido a los isótopos (p. ej. ^{12}C , ^{13}C), aductos (p. ej. $[\text{M} + \text{H}]^+$, $[\text{M} + \text{Na}]^+$) o pérdida neutra de grupos moleculares (p. ej. $[\text{M} + \text{H} - \text{H}_2\text{O}]^+$). Por ello, el preprocesado de los datos tiene como objetivo eliminar el ruido y las señales de fondo, así como agrupar en una única *feature* las diferentes señales detectadas provenientes de un mismo metabolito. Durante el preprocesado también se realiza el alineamiento de los picos cromatográficos entre las diferentes muestras. Dependiendo de la herramienta empleada, el alineamiento puede realizarse antes o después de la creación de las *features*.

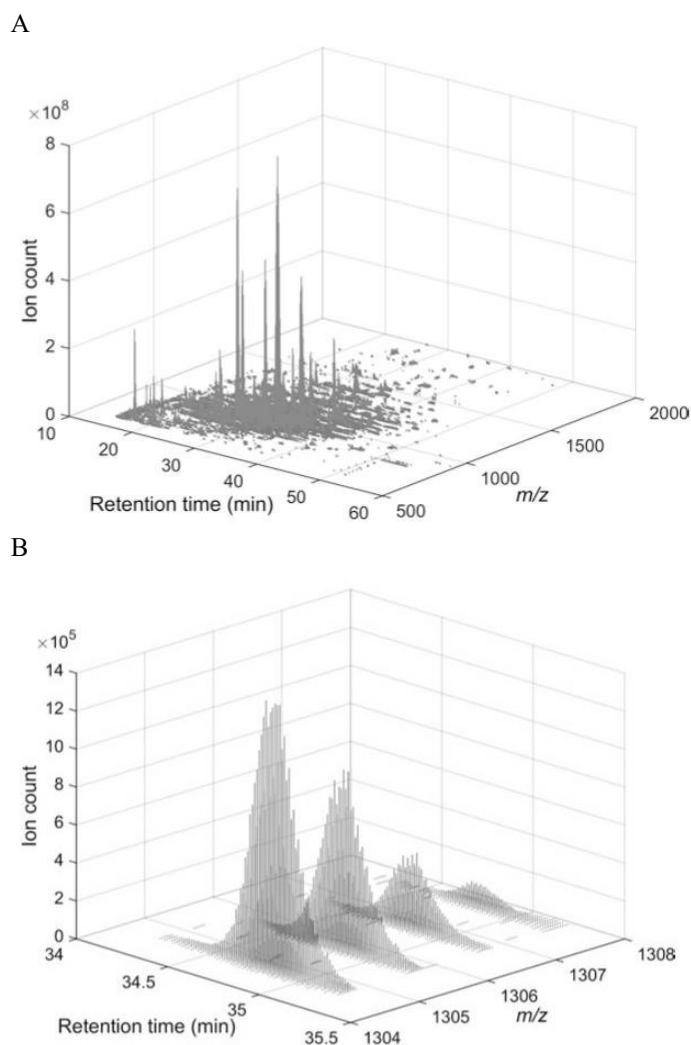


Figura 1.2.2. A) Representación 3D de todas las señales detectadas mediante LC/MS. B) Representación de señales con diferente relación m/z generadas por diferente abundancia de isótopos en un mismo metabolito. Figura tomada de [10].

Actualmente existe una gran variedad de programas, libres y comerciales, para realizar las operaciones mencionadas. Entre las herramientas libres destacan XCMS [11],

OpenMS [12] y MZmine 2 [13]. Por otro lado, algunos de los programas comerciales son MassHunter y MassProfiler Professional (Agilent Technologies, Santa Clara, CA, USA), Compound Discoverer (Thermo Fisher Scientific, Waltham, MA, USA) y Progenesis QI (Waters Corporation, Milford, MA, USA).

Tras el preprocesado se obtiene una matriz de datos que asigna a cada feature su masa monoisotópica expresada como la relación m/z , el RT y la intensidad detectada en cada una de las muestras. La matriz puede ser sometida a procesos de filtrado para eliminar las *features* ausentes en un determinado número de muestras [14]. Asimismo, es frecuente el uso de análisis estadístico univariante y multivariante con el objetivo de identificar las *features* que presentan una abundancia diferencial significativa entre los distintos grupos [15].

La identificación de los metabolitos es un proceso fundamental en los experimentos de metabolómica no dirigida basados en LC/MS con ESI como técnica de ionización. Asimismo, la identificación es el principal cuello de botella de dichos estudios debido al desconocimiento del metaboloma, el carácter incompleto de las bases de datos experimentales y la existencia de metabolitos con masas muy similares que, en el caso de los compuestos isómeros, llega a ser idéntica. La Iniciativa de Estándares en Metabolómica (en inglés *Metabolomics Standard Initiative*, MSI) establece cinco niveles en los que se puede identificar un metabolito, de mayor a menor grado de seguridad. El nivel 0 se asigna a los compuestos identificados de manera inequívoca mediante el aislamiento y la caracterización estereoquímica. El nivel 1 exige la similitud con al menos dos propiedades ortogonales (p. ej. relación m/z , RT, espectro de fragmentación) del estándar del compuesto analizado en las mismas condiciones experimentales. El nivel 2 exige la similitud con propiedades fisicoquímicas y/o espectrales de estándares analizados en otros laboratorios y cuya información está contenida en las bases de datos. El nivel 3 se alcanza mediante la comparación con propiedades fisicoquímicas y/o espectrales de clases químicas a la que pertenecen diferentes compuestos. Finalmente, el nivel 4 se corresponde con compuestos no identificados. Los niveles 0 y 1 son los más adecuados para realizar la interpretación de los resultados. Sin embargo, el número limitado de estándares químicos que cubran todo el metaboloma así como su alto coste hacen que muchos estudios metabolómicos estén basados en identificaciones de nivel 2, que en este trabajo denominaremos anotaciones putativas [7], [16].

Finalmente, tras identificar los metabolitos que presentaron una abundancia diferencial estadísticamente significativa entre los grupos comparados, se procede con la interpretación bioquímica de los resultados. En esta última etapa se pretende identificar las rutas metabólicas alteradas bajo una determinada condición (p. ej. enfermedad o tratamiento), monitorizar la eficacia de un tratamiento y encontrar nuevos biomarcadores para el diagnóstico/pronóstico de enfermedades, entre otras aplicaciones.

1.3. Motivación del Proyecto: Límites en la identificación de los metabolitos

En el caso de datos obtenidos en experimentos de metabolómica no dirigida usando LC/MS como plataforma analítica, las anotaciones putativas se obtienen utilizando programas (p. ej. XCMS, Ceu Mass Mediator 3.0, MetaboSearch) [11], [17], [18] que comparan la masa monoisotópica de las *features* con la masa de los compuestos contenidos en las bases de datos (p. ej. HMDB, Metlin, Kegg) [3], [5], [6]. Así, aunque los analizadores utilizados en estos estudios sean de alta resolución (OrbiTrap o TOF), el elevado número de metabolitos que podrían tener la misma masa monoisotópica (por ejemplo, en el caso de los isómeros) o que podrían ser el resultado de distintos tipos de aductos hace que cada *feature* pueda tener anotado varios posibles metabolitos. Tras obtener las matrices con las anotaciones putativas, el investigador debe determinar qué identificaciones son más probables para realizar la interpretación biológica de los resultados (en el caso de alcanzar el nivel 3 de las identificaciones propuesto por la MSI) o para priorizar la compra de los estándares químicos o del análisis en modalidad MS/MS (en el caso de querer alcanzar el nivel 1 o 2 de las identificaciones propuesto por la MSI, respectivamente). Con frecuencia, se priorizan aquellas anotaciones que presentan un menor error (expresado en ppm) calculado comparando la diferencia entre la masa monoisotópica experimental del analito y aquella teórica reportada en las bases de datos teniendo en cuenta los posibles aductos con los cuales los metabolitos se encuentran asociados más frecuentemente tal y como se reporta en la literatura. De esta forma, las anotaciones con menor error o que presentan un aducto más frecuente son consideradas más probables, frente a las que presentan un aducto poco común o un error más elevado. Por otro lado, los compuestos procedentes de la dieta, de fármacos o con elementos halogenados que no fueron suministrados al organismo estudiado son consideradas

identificaciones poco probables, por lo que serán descartadas en favor de otras más verosímiles.

Actualmente no existen herramientas que automaticen la simplificación de las matrices con las anotaciones putativas para facilitar la priorización de las diferentes identificaciones asignadas a una misma *feature*. Por lo tanto, los investigadores deben revisar manualmente un número muy elevado de anotaciones, comparar sus errores asociados, sus aductos y, con frecuencia, buscar en las bases de datos si el posible compuesto procede de la dieta o de un fármaco. Dado que este proceso consume mucho tiempo y esfuerzo, en el marco de este proyecto hemos desarrollado un conjunto de programas que faciliten al investigador la simplificación de las matrices y la priorización de las posibles identificaciones.

Los programas que se pretenden desarrollar en el marco de este proyecto están orientados a la manipulación y el procesamiento de grandes conjuntos de datos sin que impliquen un alto coste computacional, ni análisis estadísticos o representaciones gráficas. Por ello, los diferentes programas desarrollados fueron programados en Python, un lenguaje de programación que, junto a librerías específicas, se adapta a las necesidades de desarrollo descritas en este trabajo.

Con frecuencia, las diferentes herramientas informáticas son desarrolladas como aplicaciones de escritorio o como aplicaciones web, donde cada una presenta diferentes ventajas e inconvenientes. Las aplicaciones de escritorio son ejecutadas en el ordenador del usuario, por lo que dependen únicamente de las capacidades de procesamiento y de memoria del equipo local, lo que hace que la ejecución pueda llegar a ser muy eficiente en equipos con un alto rendimiento. No obstante, las aplicaciones de escritorio pueden generar problemas de compatibilidad y dependencias. Por otro lado, las aplicaciones web no requieren la instalación del programa en el ordenador del usuario, ni precisan ninguna dependencia más allá de poseer conexión a Internet y un navegador instalado en el equipo, lo que evita en gran medida posibles problemas de compatibilidad. Asimismo, aunque la aplicación web permite utilizar siempre la última versión del programa, esto podría ser un inconveniente en caso de que se desee utilizar una versión antigua. En el marco de este proyecto hemos integrado los diferentes programas desarrollados en una aplicación de escritorio y en una aplicación web, ambas de código abierto, lo que permitiría al usuario utilizar la opción que mejor se ajuste a sus necesidades, además de poder revisar y modificar el código siguiendo sus propios criterios.

1.4. Objetivos

El objetivo principal del proyecto consiste en desarrollar un conjunto de programas para el manejo y la simplificación de las tablas con las anotaciones putativas, lo que facilitaría la interpretación biológica de los resultados por parte del investigador.

Para la consecución del objetivo principal descrito, planteamos la realización de los siguientes objetivos parciales:

- Desarrollo de un programa para la clasificación de metabolitos de acuerdo con la información contenida en las bases de datos. La herramienta deberá identificar los metabolitos exógenos caracterizados como nutrientes, principios activos de fármacos, microbianos, halogenados o péptidos.
- Desarrollo de un programa que identifique metabolitos isómeros anotados a una misma *feature* para fusionarlos en una misma anotación. No solo se realizará esta tarea con metabolitos isómeros, sino también con aquellos compuestos que pueden aparecer en diferentes formas (p. ej. ácido y base) o que están expresados con diferente nomenclatura.
- Desarrollo de un programa que combine información de diferentes anotaciones utilizando criterios determinados por el usuario.
- Desarrollo de un programa que permita unificar en una única tabla la información que las *features* pudieran tener en diferentes matrices.
- Desarrollo de una aplicación de escritorio que integre los diversos programas desarrollados para que su utilización sea rápida e intuitiva para el usuario.
- Implementación de un servidor web que también ejecute los programas y así ampliar las opciones de utilización.

2

Diseño y Desarrollo

2.1. Introducción: Estructura del programa

El objetivo del presente Proyecto ha sido desarrollar un conjunto de programas que permitan agilizar el manejo y la simplificación de las tablas con las anotaciones putativas, facilitando la visualización de los resultados y la extracción de información útil por parte del investigador. Por ello, se ha desarrollado en el lenguaje de Python cuatro programas o módulos independientes integrados bajo una misma aplicación, lo que permite al usuario diseñar flujos de trabajo personalizados seleccionando los módulos de interés en el orden de actuación deseado. El código desarrollado para este proyecto se encuentra almacenado en GitHub (<https://github.com/CNIC-Proteomics/TurboPutative>).

A continuación, procedemos a explicar en detalle cada uno de los módulos desarrollados.

2.1.1. Tagger: Clasificación de metabolitos

Tagger es un clasificador que permite identificar metabolitos caracterizados como nutrientes, principios activos de fármacos, microbianos, péptidos o halogenados. La principal ventaja de este programa consiste en la posibilidad de clasificar a partir de la lista de las posibles identificaciones aquellas anotaciones que pertenecen a una de las cinco categorías mencionadas anteriormente. De hecho, dependiendo del experimento, dichas clasificaciones podrían considerarse como improbables, por ejemplo, en el caso de una anotación que corresponde a un fármaco que no fue suministrado al organismo estudiado y que, sin embargo, aparece en las anotaciones putativas.

Mientras la clasificación de los compuestos halogenados y peptídicos se realiza mediante el uso de expresiones regulares que permiten buscar patrones, en el caso de los nutrientes y fármacos se utilizan listas predefinidas obtenidas a partir de HMDB [3] y PubChem [19]. Las listas predefinidas se actualizarán de manera periódica para poder incorporar las modificaciones más recientes que se reportan en las bases de datos. De esta

forma se obviaría la necesidad de realizar la clasificación mediante acceso programático a HMDB y PubChem, lo que supondría la ventaja de que las bases de datos empleadas estarían actualizadas sin tener que ralentizar en gran medida el proceso, ya que la totalidad de las consultas realizadas mediante acceso programático consumiría una gran cantidad de tiempo.

A continuación, explicamos los detalles para cada una de las anotaciones posibles:

- **Nutrientes:** Para construir la lista de nutrientes utilizada por Tagger se utilizó la base de datos de HMDB [3], que incluye más de 40,000 metabolitos endógenos y exógenos detectados en humanos. Una de las principales ventajas que presenta su última versión, HMDB 4.0 [4], es el uso de una ontología estructurada para caracterizar los metabolitos. Esto permite identificar y extraer aquellos compuestos caracterizados como “*Food*” o “*Food and Nutrition*”, excluyendo los metabolitos cuyo origen podría ser endógeno (“*Endogenous*”). La base de datos de metabolitos de HMDB fue descargada de su página web (<https://hmdb.ca/downloads>) en formato XML. Se desarrollaron dos programas en el lenguaje *Bash* (`hmdb_food_extractor.sh` y `hmdb_diterpenoids_extractor.sh` en la carpeta `/tools/Tagger/dbs/scripts/`) que permiten la extracción de los metabolitos de interés y empezar a construir la lista deseada. Dado que un mismo metabolito se puede nombrar de diferentes formas, fue necesario “enriquecer” la lista de metabolitos con sus sinónimos realizando acceso programático a PubChem (`/tools/Tagger/dbs/scripts/getAllSynonyms.ipynb`). Es posible descargar el fichero con la lista de nutrientes (`/tools/Tagger/dbs/food_list.tsv`). Otra base de datos de gran relevancia en el ámbito nutricional es FooDB, que, sin embargo, no se ha podido incorporar todavía a la lista usada por Tagger. Esto se debe a que, a diferencia de HMDB, FooDB no emplea una ontología o clasificación que permita distinguir los compuestos exclusivamente exógenos de los endógenos. Siguiendo estas pautas se clasificaría como nutriente, por ejemplo, la fenilalanina, pero no la alanina.
- **Fármacos:** Para construir la lista de fármacos utilizada por Tagger se usó la base de datos de HMDB, que gracias a su ontología estructurada permite extraer los compuestos caracterizados como “*Drug*” o “*Pharmaceutical industry*”, excluyendo los compuestos cuyo origen podría ser endógeno (“*Endogenous*”).

Así, se extrajeron los metabolitos de interés de la base de datos de HMDB (/tools/Tagger/dbs/scripts/hmdb_drug_extractor.sh). Asimismo, pudimos incluir todos los sinónimos posibles provenientes de PubChem y guardarlos en un fichero (/tools/Tagger/dbs/drug_list.tsv). Una base de datos de gran relevancia en el ámbito de la farmacología es DrugBank [20]. No obstante, al igual que ocurría con FooDB, no pudimos utilizar esta base de datos debido a que no era posible distinguir de forma automática los compuestos cuyo origen procedía exclusivamente de fármacos de aquellos compuestos de origen endógeno que podrían ser productos de fármacos. Siguiendo estas pautas se clasificaría, por ejemplo, la ampicilina, pero no la epinefrina.

- **Compuestos microbianos:** Para la identificación de los compuestos de origen microbiano se utilizó una lista de elaboración propia en la que se incluyeron los sinónimos de los distintos compuestos mediante acceso programático a PubChem (/tools/Tagger/dbs/microbial_list.tsv).
- **Compuestos halogenados:** Para la identificación de los compuestos halogenados se optó por el uso de expresiones regulares. Las expresiones regulares nos permiten definir patrones con los que encontrar una determinada combinación de caracteres en una cadena de texto. En este caso, Tagger utiliza una expresión regular que, aplicada sobre el nombre de un metabolito o a su fórmula química, permite determinar la presencia de un halógeno (Tabla 2.1.1.1).

Expresión regular	([Ff]luor(?!ene) [Cc]hlor(?!ophyl) [Bb]rom [Ii]od)
Compuesto 1	14,14,14-Trifluoro-11E-tetradecenyl acetate
Compuesto 2	6-(2-Chloroallylthio)purine
Compuesto 3	Bromhexine
Compuesto 4	Fluorene
Compuesto 5	Chlorophyl

Tabla 2.1.1.1. Expresión regular utilizada por el módulo *Tagger* para identificar los compuestos con elementos halógenos. En rojo se muestra la cadena de caracteres capturada por la expresión regular. Los tres primeros compuestos serían clasificados como halógenos, a diferencia de los dos últimos. Más ejemplos en <https://regex101.com/r/Nt6A1p/1/>.

- **Compuestos peptídicos:** Al igual que los compuestos halogenados, los compuestos peptídicos son detectados por medio de expresiones regulares, que reconocen los aminoácidos en el nombre del compuesto. Para ello, los aminoácidos deben estar expresados en la nomenclatura de tres letras (Tabla 2.1.1.2).

Expresión regular	(?i)^(Ala Arg Asn Asp Cys Gln Glu Gly His Ile Leu Lys Met Phe Pro Ser Thr Trp Tyr Val [-,\s]){3,}\$
Compuesto 1	Pro Gly
Compuesto 2	Glu Ala
Compuesto 3	Gly Leu Cys
Compuesto 4	Met-Ala-Ala
Compuesto 5	Ile,Gly,Cys

Tabla 2.1.1.2. Expresión regular utilizada por el módulo *Tagger* para identificar los compuestos peptídicos. En rojo se muestra la cadena de caracteres capturada por la expresión regular. Los cinco compuestos serían clasificados como péptidos. Más ejemplos en <https://regex101.com/r/D36sC1/1>.

Cabe destacar que el usuario puede modificar algunos parámetros antes de iniciar la ejecución del programa. Por ejemplo, dependiendo de las condiciones experimentales (uso de fármacos o de compuestos halogenados, entre otras opciones), el usuario puede seleccionar cuáles clasificaciones desea realizar (las cinco o una selección de algunas de ellas) y qué expresiones regulares utilizar para identificar los metabolitos. Lo único que el usuario tiene que introducir es una tabla, en formato Excel, que debe contener al menos una columna de nombre “*Name*” y, opcionalmente, otra columna de nombre “*Chemical Formula*”, donde se encuentre el nombre de los metabolitos anotados y su fórmula química. Tras la ejecución, *Tagger* generará una tabla igual a la introducida, con una columna adicional por cada clasificación realizada (Figura 2.1.1.1).

Experimental mass	Name
102.0681	Isovalerate
162.1131	Ethyl levulinate
195.0876	L-Pinitol
274,123	Gly Leu Cys
401.2131	Galpha-Fluoroprogesterone



Experimental mass	Name	Peptide	Halogenated	Drug	Food
102.0681	Isovalerate				
162.1131	Ethyl levulinate				Food
195.0876	L-Pinitol			Drug	Food
274,123	Gly Leu Cys	Pep			
401.2131	Galpha-Fluoroprogesterone		x		

Figura 2.1.1.1. Ejemplo de la clasificación de metabolitos contenidos en una tabla empleando el módulo *Tagger* y seleccionando las opciones de clasificación *Peptide*, *Halogenated*, *Drug*, y *Food*.


2.1.2. RENAME: Eliminación de información redundante

Muchos de los metabolitos candidatos anotados a las *features* detectadas son compuestos isómeros, que tienen la misma fórmula molecular pero distinta estructura química. De hecho, en el caso de *features* obtenidas con un experimento metabolómico no dirigido que emplea LC/MS o CE/MS como plataformas analíticas, no es posible identificar, sin previo análisis en modalidad MS/MS (y en algunos casos ni siquiera con ese tipo de análisis), ni la posición ni la distribución espacial de dobles enlaces o de los grupos funcionales. Esto también sucede con compuestos que pueden aparecer reportados en la base de datos siguiendo distinta nomenclatura, como es el caso de los ácidos y sus bases conjugadas. En todos estos casos, la tabla de posibles anotaciones obtenida a partir de la masa monoisotópica de los analitos significativos utilizando programas como Ceu Mass Mediator 3.0 [17], MetaboSearch [18] o XCMS [11], resulta caracterizada por múltiples anotaciones de carácter genérico, con utilidad limitada para el investigador y con muchas repeticiones.

Por lo tanto, con el objetivo de facilitar la visualización de las tablas y la extracción de información útil por parte del investigador hemos desarrollado RENAME, un analizador capaz de identificar compuestos isómeros y equivalentes a partir de sus nombres para fusionarlos bajo una única anotación.

RENAME también utiliza expresiones regulares que permiten modificar el nombre de los compuestos para eliminar la información relativa a la isomería (p. ej. posición de grupos funcionales, dobles enlaces y quiralidad, entre otros). En el caso de compuestos que no son isómeros pero que pueden tener distintas formas, como el ácido y la base, se convierten todos los compuestos a una de sus formas (p. ej. convertir todos los ácidos a bases). Tras aplicar las expresiones regulares, RENAME identifica los metabolitos que deben ser fusionados bajo una única anotación comparando los nombres obtenidos tras el procesado (Figura 2.1.2.1).

Experimental mass	Name
131.0582	N-acetyl-L-Alanine
131.0582	N-Acetyl-beta-alanine
217.1045	(2R,3R)-heptane-1,2,3-triol
217.1045	heptane-1,2,3-triol
311.1262	6E,8E,14E-Hexadecatriene-10,12-diynoic acid
311.1262	6E,8E,14Z-Hexadecatriene-10,12-diynoic acid



Experimental mass	Name
131.0582	acetyl-Alanine
217.1045	heptane-triol
311.1262	Hexadecatriene-diynoate

Figura 2.1.2.1. Procesado del nombre de metabolitos utilizando el sistema de expresiones regulares empleado por el módulo *RENAME* (/tools/RENAME/regex.ini).

Las expresiones regulares usadas se encuentran en un fichero (/tools/RENAME/regex.ini) dividido en secciones con diferentes parámetros. Cada una de las secciones suponen una unidad de procesamiento aplicado sobre el nombre de los compuestos. Así, la sección tiene dos parámetros con distinta función. El primero de ellos contiene la expresión regular que permite detectar una combinación determinada de caracteres en el nombre del compuesto. El segundo parámetro contiene los caracteres que

reemplazarán a la cadena detectada por la expresión regular. Cada uno de estos procesamiento se aplican de manera secuencial sobre el nombre de los compuestos (Figura 2.1.2.2).

Nombre de entrada			Hexadecatriene-10,12-diynoic acid
Sección 1	Regex	<code>([-\s])(\d+[\s]{2})+-</code>	Hexadecatriene-10,12-diynoic acid
	Replace	<code>\g<1></code>	Hexadecatriene-diynoic acid
Sección 2	Regex	<code>(?i)ic acid</code>	Hexadecatriene-diynoic acid
	Replace	<code>ate</code>	Hexadecatriene-diynoate
Nombre de salida			Hexadecatriene-diynoate





Figura 2.1.2.2. Mecánica del procesamiento de los nombres de los metabolitos con las expresiones regulares.

No obstante, hay compuestos que, además de ser procesados con las expresiones regulares, reciben un tratamiento adicional por parte de RENAME. Es el caso de los péptidos y de los lípidos que presentan ácidos grasos en parte de su estructura (p. ej. fosfolípidos, esfingolípidos y glicerolípidos).

En el primer caso, RENAME combina bajo una única anotación aquellos péptidos que tienen la misma composición aminoacídica, aunque con distinta secuencia. Así, únicamente se mantendrá una de las secuencias comparadas. Para ello, los aminoácidos deben estar expresados en la nomenclatura de tres letras (Figura 2.1.2.3).

Experimental mass	Name
388.1555	Pro Pro Met
388.1555	Pro Met Pro
388.1555	Met Pro Pro
433.1408	Trp Asp Asp
433.1408	Asp Asp Trp
433.1408	Asp Trp Asp



Experimental mass	Name
388.1555	Met Pro Pro
433.1408	Trp Asp Asp

Figura 2.1.2.3. Combinación de péptidos con la misma composición aminoacídica usando el módulo *REname*.

En el caso de los lípidos con ácidos grasos, el procesamiento adicional se realiza utilizando el paquete Goslin [21]. Goslin puede procesar el nombre de lípidos provenientes de diferentes bases de datos (p. ej. HMDB [3], LipidMaps [22], SwissLipids [23]) y clasificarlo gracias a un sistema de analizadores y gramáticas predefinidas. La información extraída del nombre del lípido se almacena en un *objeto* de estructura fija, que facilita al programador el acceso a la información de interés. En este Proyecto estamos interesados en extraer la siguiente información de los compuestos lipídicos:

- Grupo principal (p. ej. PE o fosfoetanolamina), también denominado *head group*.
- Número total de átomos de carbono en los ácidos grasos.
- Número total de dobles enlaces en los ácidos grasos.
- Tipo de enlace (éter o vinil éter) de los ácidos grasos.
- Número de grupos hidroxilo y metilo en los ácidos grasos.

Tras el procesado, *REname* puede fusionar aquellos lípidos isómeros comparando los nombres obtenidos (Figura 2.1.2.4).


Experimental mass	Name		Experimental mass	Name
793.5565	PE-Cer(d16:2(4E,6E)/24:0(2OH))		793.5565	PE-Cer(40:2(2OH))
800.5444	PE(18:0(10(R)Me)/16:0)		800.5444	PE(34:0(Me))
812.5478	PC(15:0/18:1(9Z))		812.5478	PC(33:1)
812.5478	PC(15:1(9Z)/18:0)		869.5545	PI(38:3)
869.5545	PI(18:3(6Z,9Z,12Z)/20:0)			
869.5545	PI(20:2(11Z,14Z)/18:1(9Z))			

Figura 2.1.2.4. Procesado del nombre de lípidos isómeros con ácidos grasos usando el módulo *REname* y el paquete Goslin.

En lo que se refiere a su ejecución, *REname* debe recibir como entrada una tabla en formato Excel. Esta debe tener al menos dos columnas con la masa de las *features* extraídas (*Experimental mass*) y el nombre de los compuestos anotados (*Name*). Asimismo, el usuario puede modificar algunos parámetros del módulo y alterar de forma considerable el procesado del nombre de los compuestos. Por ejemplo, es posible cambiar el sistema de expresiones regulares utilizado por *REname*.

2.1.3. RowMerger: Combinación de anotaciones

RowMerger es un comparador de entidades que permite combinar la información de diferentes anotaciones según los criterios determinados por el usuario. Por lo tanto, este módulo permite agrupar anotaciones asignadas a una misma *feature* en una única entrada de la tabla con las anotaciones putativas, facilitando así la visualización de los datos de cara a la interpretación biológica de los resultados.

Para la ejecución de *RowMerger* el usuario debe especificar qué propiedades de las anotaciones (*i.e.* columnas de la tabla) se tendrán en cuenta para que se realice la fusión. Asimismo, deberá indicar qué propiedades se conservarán en la entrada resultante de la fusión de las anotaciones. Por defecto, *RowMerger* combinará las anotaciones que tengan la misma masa experimental (*Experimental mass*), aducto (*Adduct*), error (*mz Error (ppm)*) y peso molecular (*Molecular Weight*), de tal forma que la entrada resultante

conservará el identificador (*Identifier*) y el nombre (*Name*) de todas las anotaciones fusionadas (Figura 2.1.3.1).

RowMerger recibe como entrada una tabla, en formato Excel, y genera como salida una tabla procesada con la misma estructura.

Experimental mass	Identifier	Adduct	Name
83.0607	154396	M+H	Methylimidazole
83.0607	61184	M+H	Fomepizole
171.0628	150549	M+Na	Paratose
171.0628	158202	M+Na	Abequose
171.0628	151197	M+Na	Tyvelose



Experimental mass	Identifier	Adduct	Name
83.0607	154396 // 61184	M+H	Methylimidazole // Fomepizole
171.0628	150549 // 158202 // 151197	M+Na	Paratose // Abequose // Tyvelose

Figura 2.1.3.1. Combinación de anotaciones usando el módulo *RowMerger* para fusionar las entradas que tienen en común las propiedades *Experimental mass* y *Adduct*, conservando el valor de las propiedades *Identifier* y *Name*.

2.1.4. TableMerger: Combinación de matrices

TableMerger es un comparador de tablas que permite unificar en una única matriz la información que las *features* pudieran tener en diferentes conjuntos de datos. Así, muchas de las herramientas utilizadas para realizar las anotaciones putativas (p. ej. Ceu Mass Mediator 3.0 [17] o MetaboSearch [18]) reciben únicamente una lista con la masa experimental de las *features* detectadas, y generan una tabla con la información de los compuestos anotados. Con TableMerger es posible combinar en una única matriz la tabla con las anotaciones putativas y la información adicional (p. ej. área del pico, resultados de análisis estadísticos, RT, entre otros) de las *features* que el usuario pudiera tener en otras tablas.

Por lo tanto, TableMerger recibe como entrada dos tablas especificadas por el usuario. Las *features* equivalentes de ambas tablas son identificadas gracias a la masa

experimental (*Experimental mass*) y al tiempo de retención (*RT [min]*). No obstante, si el RT no está contenido en alguna de las tablas de entrada, se utilizará únicamente la masa experimental. La salida de este módulo será una única matriz con la información combinada de los dos conjuntos de datos (Figura 2.1.4.1).

Dado que la masa experimental de las *features* puede estar expresada con distintas cifras significativas en ambas tablas, el usuario podrá especificar el número de decimales al que realizar el redondeo. Ejemplo del funcionamiento de dicho modulo se representa en la figura 2.1.4.1.

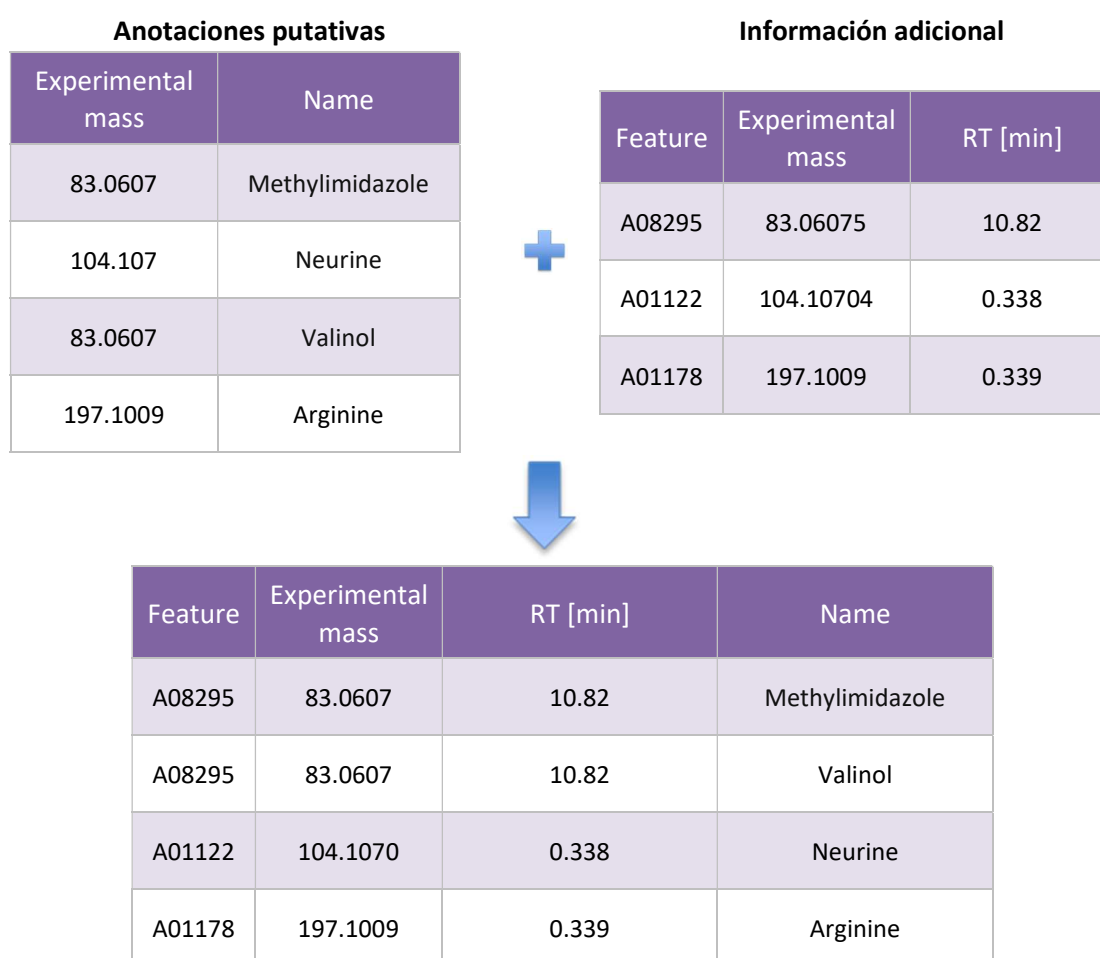


Figura 2.1.4.1. Combinación de la información de los *features* en una única tabla usando *TableMerger*.

2.2. Integración de los módulos: Aplicación de escritorio

Otro de los objetivos del Proyecto es integrar los cuatro programas desarrollados y explicados anteriormente en una aplicación de escritorio fácilmente utilizable por

expertos en el campo de la Metabolómica. Cabe destacar que el desarrollo de aplicaciones de escritorio presenta ventajas, tales como la ejecución local del programa sin necesidad de disponer de conexión a Internet, o la posibilidad de introducir modificaciones *ad hoc* en el código del programa para adaptarlo a las necesidades del usuario. Para ello, hemos utilizado los entornos de trabajo Electron [22] y Node.js [23], que permiten desarrollar aplicaciones de escritorio utilizando tecnologías y lenguajes orientados al desarrollo web (HTML, CSS y JavaScript).

La aplicación de escritorio fue desarrollada para Windows 10 y Ubuntu 18.0 con una arquitectura de 64 bits en ambos casos. Para poder utilizar la versión no compilada es necesario que el usuario tenga instalado Python 3.6 o una versión superior [24]. Así, la primera vez que se ejecute TurboPutative se creará un entorno virtual de Python [25] en el que se instalarán las librerías requeridas por los distintos módulos. No obstante, también es posible descargar una versión que, pese a ser de mayor tamaño, contiene los módulos compilados, por lo que no requiere la instalación de Python.

El código y las diferentes versiones de la aplicación se pueden descargar desde la plataforma GitHub (<https://github.com/CNIC-Proteomics/TurboPutative>). Asimismo, en la aplicación es posible encontrar una documentación en la que se describe la función y los parámetros de cada uno de los módulos integrados (Figuras 2.2.1)

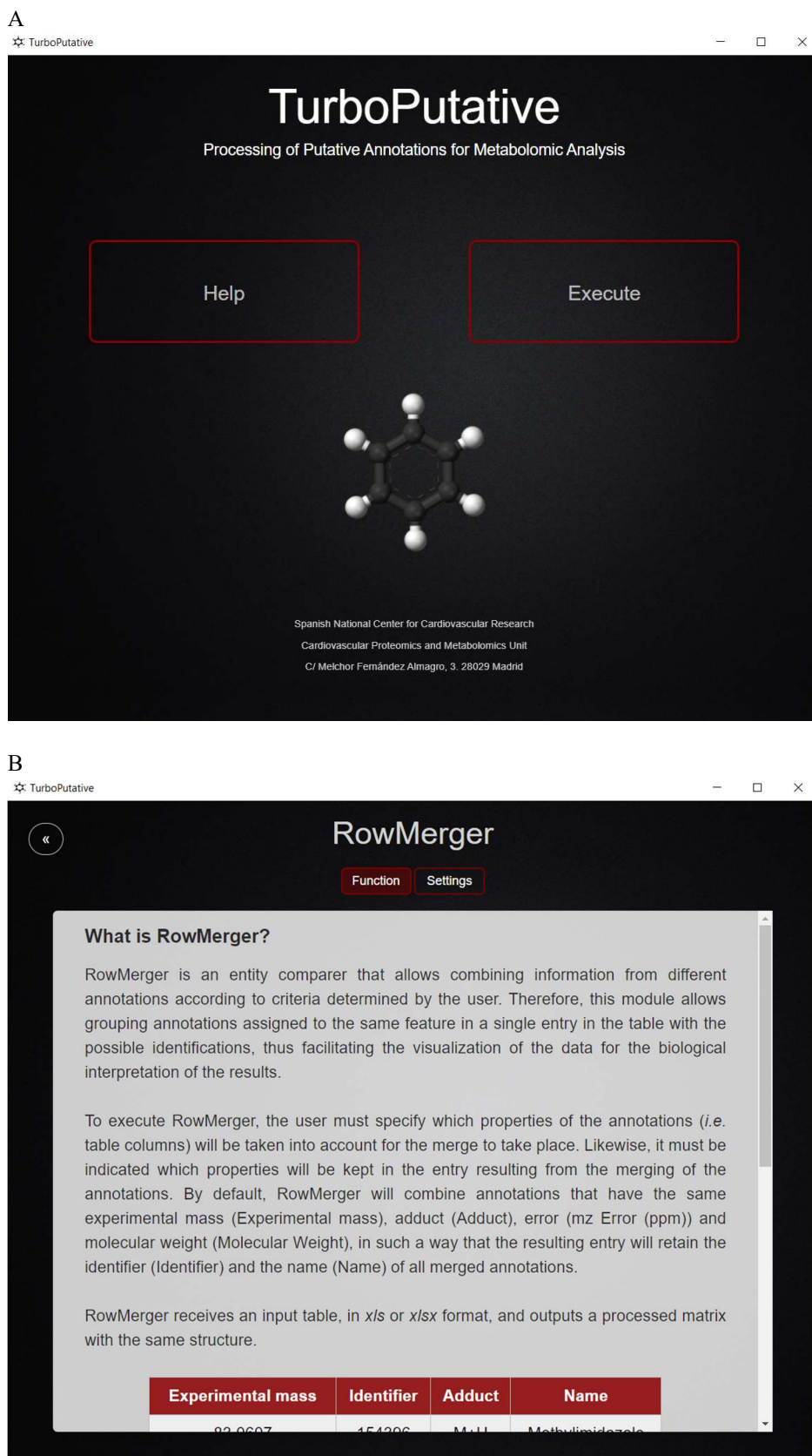


Figura 2.2.1 Interfaz de la aplicación de escritorio de TurboPutative A) Menú principal de la aplicación. B) Sección de la documentación en la que se describe la función del módulo RowMerger.

La interfaz que aparece tras seleccionar la opción *Execute* muestra los cuatro módulos. Así, el usuario debe seleccionar los módulos en el orden de ejecución deseado, de tal forma que el *output* de cada módulo será el *input* del siguiente. A continuación, el usuario debe indicar el fichero con la tabla de anotaciones putativas que desea procesar. Finalmente, deberá configurar los parámetros correspondientes a cada uno de los módulos (Figura 2.2.2). Tras iniciar la ejecución, la aplicación mostrará una pantalla de espera, que proporciona información sobre la fase en la que se encuentra la ejecución. Una vez finalizado el flujo de trabajo, el usuario podrá acceder a la carpeta con las tablas generadas por cada uno de los módulos.

En caso de producirse un error en la ejecución, la aplicación lo notificará. Así, existe un conjunto de errores “controlados” por la aplicación, en cuyo caso se indicará cuál es la causa y qué podría hacer el usuario para corregirlo (p. ej. formato de tabla, nombre de columnas, formato del fichero con expresiones regulares, entre otros). La lista de errores controlados por TurboPutative se encuentran registrados en un fichero (/tools/errorCode.json).

Tras iniciar el flujo de trabajo, la interfaz almacena los parámetros seleccionados por el usuario en diferentes ficheros. A continuación, se ejecuta la secuencia de comandos *integrator* (Batch en Windows o SH en Unix, localizado en la ruta de carpetas /desktopApp/app/src/) que, a su vez, ejecutará secuencialmente los programas en Python en el orden seleccionado.

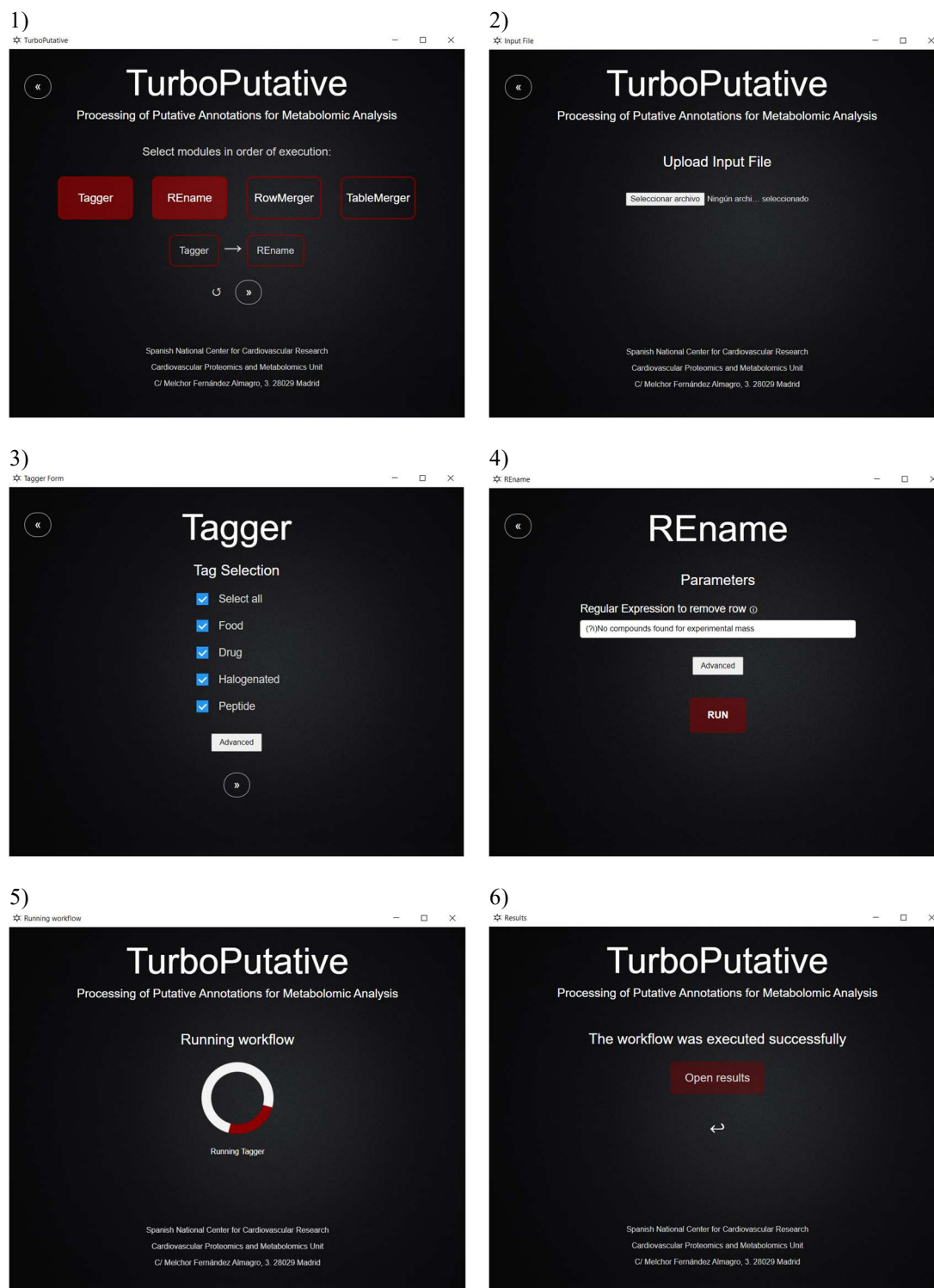


Figura 2.2.2. Diseño y ejecución del flujo de trabajo utilizando aplicación de escritorio de TurboPutative.

2.3. Integración de los módulos: Aplicación web

A pesar de sus ventajas, el desarrollo de la aplicación de escritorio conlleva algunos inconvenientes, tales como las posibles incompatibilidades causadas por una diferente versión del Sistema Operativo, o la necesidad de descargar el programa completo cada vez que se incorporan nuevas modificaciones. Por ello, con el objetivo de superar estas limitaciones, hemos integrado los módulos en una aplicación web (<https://turboputative.herokuapp.com>), que, además de ofrecer al usuario un acceso más inmediato, elimina los problemas de compatibilidades al exigir únicamente conexión a Internet.

La aplicación se encuentra alojada en un servidor gratuito de Heroku (<https://www.heroku.com/>). Por este motivo, los recursos computacionales disponibles son escasos, de tal manera que la ejecución del flujo de trabajo se puede demorar.

El sitio web desarrollado organiza su contenido en diferentes secciones, a las que es posible acceder a través de un menú de navegación situado en la cabecera de las diferentes páginas (Figura 2.3.1). Las secciones accesibles son:

- *Home*: Es la página principal de la web, que ofrece una sucinta descripción del programa y un llamativo acceso (“*Try it!*”) a la sección *Execute*, desde la que es posible diseñar y ejecutar el flujo de trabajo.

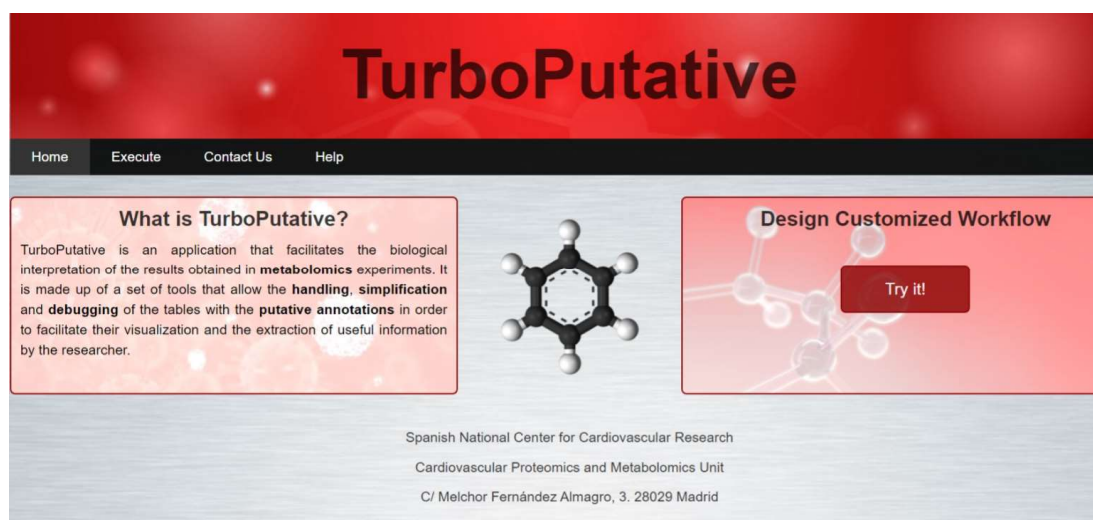
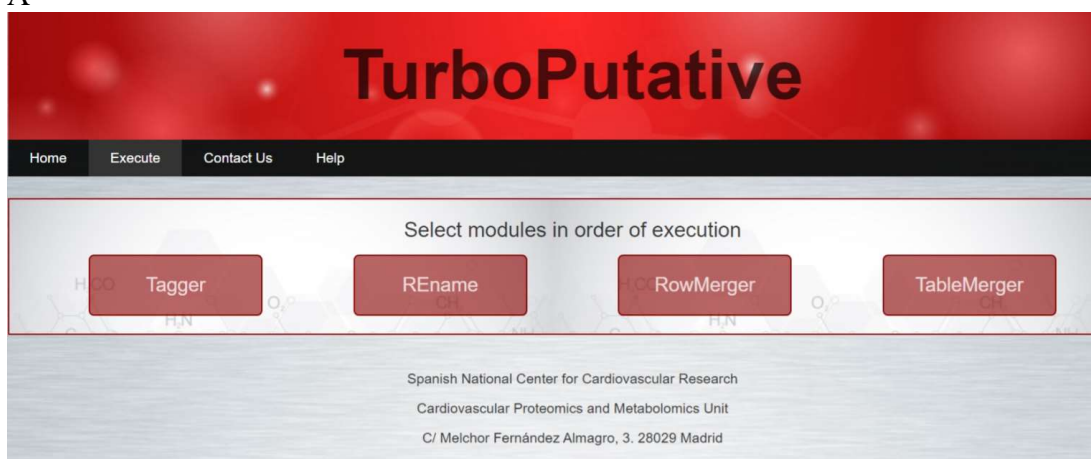


Figura 2.3.1. Pantalla principal del sitio web, que permite acceder a las diferentes secciones mediante un menú de navegación.

- *Help*: La sección *Help* se divide en dos subapartados: *Modules* y *Settings*. *Modules* contiene una detallada descripción de la función de los módulos integrantes del programa, mientras que *Settings* describe los parámetros modificables por el usuario en cada uno de los módulos.
- *Contact us*: Esta sección ofrece un mail de contacto al que enviar los posibles errores, las dudas o propuestas de mejora que los usuarios pudieran tener.
- *Execute*: La sección *Execute* permite diseñar y ejecutar un flujo de trabajo personalizado (Figura 2.3.2 y 2.3.3). A continuación, explicamos con mayor grado de detalle la estructura y funcionamiento de esta sección.

A



B

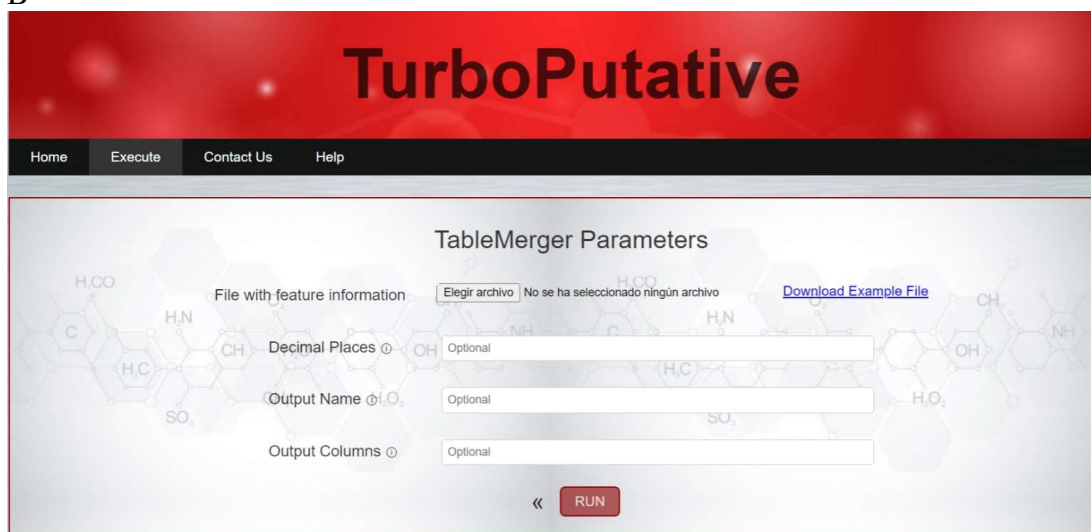
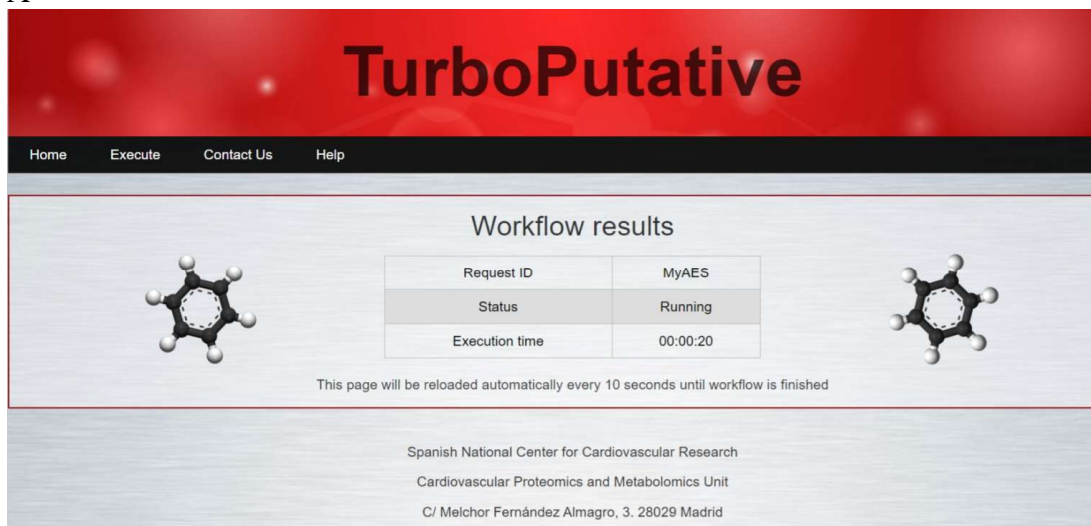


Figura 2.3.2. Sección *Execute* del sitio web, que ofrece al usuario la posibilidad de seleccionar los módulos en el orden deseado (A), y de modificar los parámetros en cada uno de ellos (B).

La primera acción que debe realizar el usuario en la sección *Execute* es la selección de los módulos en el orden de actuación deseado. A continuación, se mostrarán los parámetros configurables de las herramientas seleccionadas, que podrán ser modificados antes de la ejecución del flujo de trabajo.

A

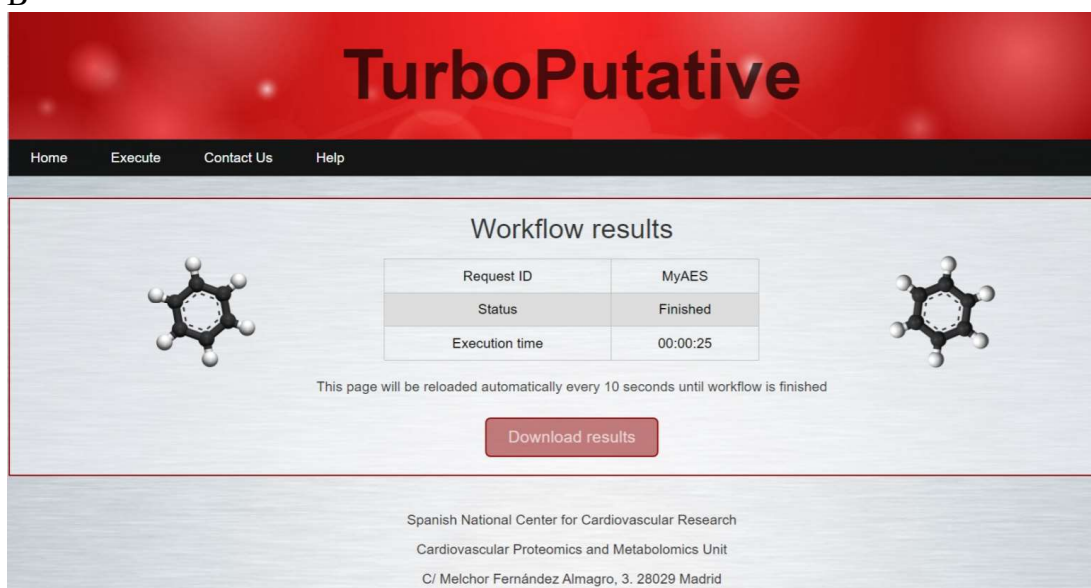


The screenshot shows the TurboPutative web interface. At the top is a red header with the title "TurboPutative". Below it is a black navigation bar with links: Home, Execute, Contact Us, and Help. The main content area has a light gray background. In the center, under the heading "Workflow results", there is a table with the following data:

Request ID	MyAES
Status	Running
Execution time	00:00:20

Below the table, a message states: "This page will be reloaded automatically every 10 seconds until workflow is finished". At the bottom, the footer text reads: "Spanish National Center for Cardiovascular Research", "Cardiovascular Proteomics and Metabolomics Unit", and "C/ Melchor Fernández Almagro, 3. 28029 Madrid".

B



The screenshot shows the TurboPutative web interface with the workflow completed. The status in the table is now "Finished". A new "Download results" button has appeared below the table. The rest of the interface, including the header, navigation bar, and footer, remains the same as in screenshot A.

Figura 2.3.3. A) Página de espera enviada por el servidor tras la ejecución del flujo de trabajo. B) Página desde la que descargar los resultados obtenidos.

El envío del trabajo al servidor se realiza mediante una petición HTTP con método POST, que contiene la tabla con las anotaciones putativas aportada por el usuario y la información de los diferentes parámetros. El servidor procesa la petición realizando dos operaciones independientes. Por un lado, se inicia el flujo de trabajo mediante la ejecución

de un programa en Bash (/webApp/src/Tools/integrator.sh), que, a su vez, ejecutará los módulos del programa en el orden apropiado. Por otro lado, el servidor envía al navegador una página de espera con el identificador del proyecto (Figura 2.3.3). Así, la página de espera podrá enviar al servidor peticiones GET cada diez segundos, que contienen en la URL dicho identificador. De esta forma, el servidor puede comprobar si el flujo de trabajo ha finalizado. En ese caso, el servidor envía al navegador una página desde la que es posible descargar los resultados en un fichero con formato de compresión ZIP (Figura 2.3.4).

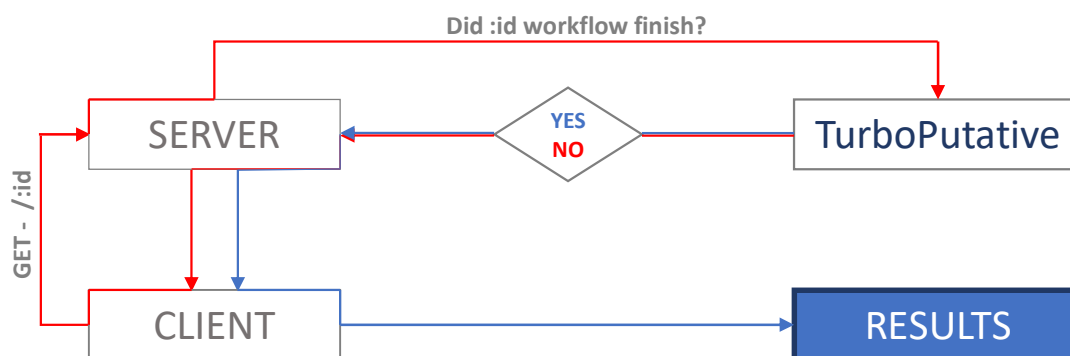


Figura 2.3.4. Diagrama de las consultas realizadas en la página de espera del sitio web. La página de espera envía peticiones GET con el identificador del trabajo al servidor. Este comprobará si dicho trabajo ha finalizado. De ser así (azul), se enviará al navegador una página desde la que descargar los resultados. En caso contrario, se devolverá nuevamente la página de espera.

Pruebas Realizadas y Resultados Obtenidos

3.1. Introducción: Obtención de la matriz con las anotaciones putativas

Con el objetivo de mostrar el funcionamiento de TurboPutative, se procesó para este proyecto un conjunto de datos de prueba correspondiente a un experimento de metabolómica no dirigida que usó como método analítico LC/ESI-MS. El estudio está enfocado en la detección de metabolitos de origen microbiano asociados con el desarrollo de la aterosclerosis. Para ello, el experimento se realizó con muestras de plasma de ratón con deficiencia de ApoE sometidos a una dieta alta en grasa y colesterol durante 8 semanas (modelo experimental de aterosclerosis). Asimismo, la dieta fue suplementada con colina, cuyo metabolito N-óxido de trimetilamina, generado por la acción de la microbiota y el hígado, aparece asociado al desarrollo de esta enfermedad [26], [27]. Para estudiar el efecto de los metabolitos de origen microbiano, los ratones fueron sometidos o no a un cóctel de antibióticos de amplio espectro disuelto en agua potable, que elimina en gran medida la microbiota intestinal. Los datos incluidos en el presente proyecto se obtuvieron a partir de aquellas *features* que, en el experimento realizado, mostraron una abundancia diferencial estadísticamente significativa entre los grupos comparados (ratones ApoE tratados con y sin antibióticos) y una correlación significativa con la placa aterosclerótica que fue medida mediante la cuantificación del tamaño de la lesión al sacrificio tras 8 semanas de dieta. Las 25 *features* de interés se reportaron en una tabla con tres columnas, que incluyen el identificador (un código alfanumérico único por cada feature), la masa monoisotópica experimental expresada como relación m/z y el tiempo de retención (/test/Feature_Data.xlsx). Las anotaciones putativas se realizaron con la herramienta “BATCH SEARCH” de Ceu Mass Mediator 3.0 (CMM) [17], para lo cual se introdujo en el campo *Experimental mass* una lista con las relaciones m/z de las *features* contenidas en la matriz inicial. Los parámetros utilizados para realizar la búsqueda se especifican en la Tabla 3.1.1 y dependen de las características analíticas experimentales (fases móviles empleadas, modalidad de ionización, entre otras).

Tolerance	10 ppm
Databases	All except MINE
Metabolites	All except peptides
Input Masses Mode	m/z masses
Ionization Mode	Positive Mode
Adducts	<ul style="list-style-type: none">- M+H- M+Na- M+K- M+H-H₂O- M+H+HCOONa

Tabla 3.1.1. Parámetros utilizados para realizar la búsqueda de posibles identificaciones en CMM.

A partir de las 25 *features* de interés, CMM generó una matriz con 540 filas, cada una de las cuales se corresponde con una posible identificación asociada a una *feature* concreta. Estos resultados fueron exportados a una tabla en formato Excel, que fue el archivo de entrada de TurboPutative (/test/MS_experiment.xls).

3.2. Simplificación automática de la matriz usando TurboPutative

Con el objetivo de mejorar la visualización y el manejo de las tablas con las posibles identificaciones, se usaron los cuatro módulos de TurboPutative, que permiten simplificar automáticamente las matrices con las anotaciones putativas y que, en última instancia, facilitan la subsecuente priorización de las distintas posibles identificaciones de los metabolitos de interés.

El primer módulo utilizado fue Tagger, seleccionando las cinco clasificaciones disponibles. Tagger recibió la matriz exportada desde CMM y generó una tabla (/test/1_tagged_MS_experiment.xlsx) similar con 5 columnas adicionales, que indicaban qué compuestos eran halogenados, peptídicos o metabolitos exógenos procedentes de la dieta, fármacos o microorganismos, por un total de 32 posibles identificaciones anotadas (Apéndice A.1).

El siguiente módulo utilizado fue REname, que permitió combinar en una única entrada las anotaciones de metabolitos isómeros o con distinta nomenclatura asignados a una misma *feature*. La tabla resultante (/test/2_REnamed_MS_experiment.xlsx) contenía

134 entradas, lo que supone una reducción del 25% respecto al tamaño de la matriz inicial y una mejora en la visualización de los resultados por parte del usuario (Apéndice A.2).

A continuación, se utilizó RowMerger con el objetivo de combinar en una única entrada la información de las identificaciones que estaban asignadas a una misma *feature* y que tenían el mismo error y aducto asociados, manteniendo el nombre y los identificadores de las anotaciones fusionadas. La tabla generada (3_rowMerged_MS_experiment.xlsx) tenía un total de 60 entradas, lo que suponía una reducción de aproximadamente el 90% respecto al número de filas de la matriz original (Apéndice A.3).

Finalmente, se utilizó TableMerger para incorporar a la tabla definitiva la información adicional contenida en la matriz inicial (identificador de la *feature* y el tiempo de retención). La tabla final que se generó (/test/4_tableMerged_MS_experiment.xlsx) tenía un total de 67 entradas, 7 de las cuales se correspondían con *features* que no pudieron ser identificadas por CMM y que se clasificaron como *unknowns* (Apéndice A.4).

Por lo tanto, cabe destacar que el uso de los diferentes módulos de TurboPutative, que se han desarrollado en el presente trabajo, ha permitido reducir considerablemente las dimensiones de la matriz inicial (Figura 3.2.1). Concretamente, a partir de las 540 filas iniciales resultante de la búsqueda en base de datos (CMM, en este caso), el programa ha conseguido reducir la tabla a 60 entradas, generando una tabla final más visual y

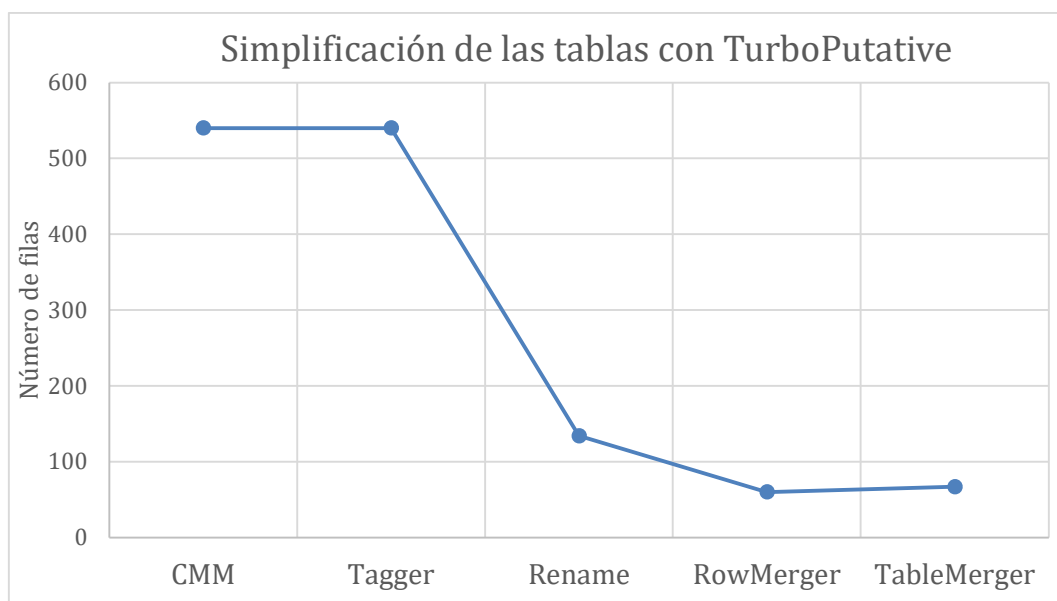


Figura 3.2.1. Reducción del número de filas de las tablas con las anotaciones putativas usando los cuatro módulos de TurboPutative.

manejable en la que es posible realizar la priorización y descarte de las posibles identificaciones de los metabolitos de una forma más eficiente e intuitiva por parte del usuario/investigador.

Las clasificaciones realizadas por Tagger muestran, por ejemplo, que los metabolitos candidatos “*5-chloropentanoic acid*” y “*Honaucin A*”, asignados a la *feature* A01654, y “*2-Chlorobiphenyl*” y “*4-Chlorobiphenyl*”, asignados a la *feature* A01556, son compuestos halogenados, y que, al no haber sido suministrados a los organismos estudiados, pueden ser considerados por el usuario como identificaciones poco probables. Asimismo, el metabolito candidato “*Trimethylamine-N-oxide*”, asignado a la *feature* A01283, cobra una especial relevancia en este estudio al haber sido clasificado como de origen microbiano.

Por otro lado, las simplificaciones realizadas por RENAME y RowMerger permiten comprobar, por ejemplo, que entre los metabolitos candidatos asignados a la *feature* A01917 encontramos diferentes monosacáridos como los isómeros de la desoxirribosa, la desoxixilulosa o la desoxiarabinosa. Las diversas anotaciones de estos compuestos fueron combinadas en una única entrada de la tabla final, facilitando en gran medida su visualización. De un modo similar es posible observar que la *feature* A05006 probablemente se corresponda con ácidos biliares (p. ej. “*Taurocholic acid*”, “*Taurohyocholate*” o “*Tauro-b-muricholic acid*”, entre otros) o que la *feature* A02242 posiblemente se corresponda con compuestos derivados de monosacáridos, como la fructosamina o los diferentes isómeros de la manosamina.

Cabe destacar que todo este proceso supone un importante ahorro de tiempo por parte del usuario, sobre todo si consideramos que la ejecución de TurboPutative requirió un total de 30 segundos de tiempo de CPU en un solo hilo para procesar la matriz exportada desde CMM en lugar de las múltiples búsquedas en bases de datos que se realizarían por el usuario *a posteriori* y de forma manual para obtener una simplificación de las tablas comparable a aquella presentada en este trabajo.

Finalmente, tras determinar los metabolitos candidatos más probables en cada una de las *features*, el investigador puede proceder de distintas formas. Así, es posible continuar con la interpretación biológica de los resultados, o bien realizar análisis adicionales para confirmar las posibles identificaciones. En este sentido, se podría utilizar la información extraída de las tablas para priorizar la compra de estándares sintéticos o los análisis en modalidad MS/MS, que permitirían alcanzar niveles de identificación 1 y

2, respectivamente, de acuerdo con los criterios establecidos por la MSI. Más detalles sobre las posibles identificaciones de los metabolitos incluidos en las tablas finales que se reportan en el Apéndice A del presente trabajo.

4

Conclusiones y Trabajo Futuro

4.1. Conclusiones

A continuación, procedemos a citar las conclusiones que se derivan de la realización de este trabajo:

- Las herramientas desarrolladas en este proyecto permiten simplificar de manera automática las matrices con las anotaciones putativas. Tagger es capaz de identificar los compuestos halogenados y peptídicos, así como muchos de los metabolitos exógenos procedentes de la dieta, fármacos y microorganismos. RENAME permite combinar las anotaciones de metabolitos isómeros, o de compuestos expresados con distinta nomenclatura. RowMerger fusiona en una única anotación las identificaciones asignadas a una misma *feature* que tienen el mismo error y aducto asociados, conservando el nombre de todos los metabolitos candidatos. TableMerger permite unificar en una única tabla la información que las *features* pudieran tener en diferentes conjuntos de datos. Las tablas generadas por estas herramientas pueden ser revisadas por el investigador de manera más eficiente e intuitiva, lo que permite realizar la priorización y descarte de las posibles identificaciones de los metabolitos candidatos más rápidamente.
- Los programas desarrollados fueron integrados en una aplicación de escritorio de código abierto que puede ser utilizada por el usuario de manera rápida e intuitiva. Así, el usuario puede introducir modificaciones *ad hoc* en el código de la aplicación, en las listas de compuestos y en las expresiones regulares utilizadas por los distintos módulos para adaptar el funcionamiento del programa a sus necesidades.
- Además, se desarrolló una aplicación web, que ofrece al usuario un acceso más inmediato a las funciones del programa evitando los posibles problemas de

dependencia o incompatibilidad que pudieran producirse en la aplicación de escritorio.

4.2. Trabajo Futuro

Finalmente, mostramos algunos de los aspectos de TurboPutative que podrían ser objeto de mejora en futuras actualizaciones:

- Ampliar las listas utilizadas por Tagger para identificar los metabolitos exógenos procedentes de la dieta, fármacos y microorganismos. Para ello, se podrían incorporar la información contenida en otras bases de datos, como FooDB o DrugBank. Ante la falta en ellas de una ontología estructurada que permita extraer únicamente los compuestos exógenos se podrían utilizar técnicas de minería de texto y aprendizaje automático que permitieran clasificar los metabolitos a partir de la descripción presente en la base de datos. Asimismo, se podrían incluir nuevas listas que facilitasen la exclusión de otros metabolitos exógenos, como por ejemplo los procedentes de plantas.
- Mejorar la capacidad de RENAME para identificar compuestos expresados con diferente nomenclatura y optimizar el rendimiento del procesado de los lípidos con el paquete Goslin. Esta última operación constituye el principal cuello de botella de TurboPutative al tener un mayor coste computacional, por lo que se podría mejorar programando en C++ la parte del código dedicada a esta tarea.

Glosario de acrónimos y abreviaturas

- **CE/MS:** Plataforma analítica empleada en metabolómica que usa la técnica de electroforesis capilar acoplada a espectrometría de masa para la separación y detección de metabolitos.
- **ESI:** Técnica de ionización en la que cationes o aniones disueltos se transfieren a la fase gaseosa mediante la formación y desolvatación a presión atmosférica de una nube de gotas cargadas que resultan de aplicar una diferencia de potencial entre la punta del capilar que contiene la solución y un contraelectrodo.
- **GC/MS:** Plataforma analítica empleada en metabolómica que usa la técnica de cromatografía de gases acoplada a espectrometría de masa para la separación y detección de metabolitos.
- **LC/MS:** Plataforma analítica empleada en metabolómica que usa la técnica de cromatografía de líquidos acoplada a espectrometría de masa para la separación y detección de metabolitos.
- **RT:** Tiempo que transcurre entre la inyección de la muestra y la aparición de la respuesta máxima en técnicas de cromatografía líquida y cromatografía de gases.
- **HMDB:** La base de datos del metaboloma humano (en inglés *Human Metabolome Database*, HMDB) es una base de datos de libre acceso que contiene información detallada sobre los metabolitos que se encuentran en el cuerpo humano.

Bibliografia

- [1] H. H. Marco Fernandes, Bela Sanches, “Cheminformatics and Computational Approaches in Metabolomics,” in *Computational Biology [Internet]*, E. Husi H, Ed. Brisbane (AU): Codon Publications, 2019.
- [2] G. S. Carlos Guijas, J Rafael Montenegro-Burke, Benedikt Warth, Mary E Spilker, “Perspective Metabolomics activity screening for identifying metabolites that modulate phenotype,” *Nat. Biotechnol.*, vol. 36, no. 4, pp. 316–320, 2018.
- [3] L. Q. David S Wishart, Dan Tzur, Craig Knox, Roman Eisner, An Chi Guo, Nelson Young, Dean Cheng, Kevin Jewell, David Arndt, Summit Sawhney, Chris Fung, Lisa Nikolai, Mike Lewis, Marie-Aude Coutouly, Ian Forsythe, Peter Tang, Savita Shrivastava, Kevin Jeroncic, , “HMDB : the Human Metabolome Database,” *Nucleic Acids Res.*, vol. 35, no. Database issue, pp. 521–526, 2007.
- [4] A. S. David S Wishart, Yannick Djoumbou Feunang, Ana Marcu, An Chi Guo, Kevin Liang, Rosa Vázquez-Fresno, Tanvir Sajed, Daniel Johnson, Carin Li, Naama Karu, Zinat Sayeeda, Elvis Lo, Nazanin Assempour, Mark Berjanskii, Sandeep Singhal, David Arndt, Yonjie Liang, “HMDB 4.0 : the human metabolome database for 2018,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. 608–617, 2018.
- [5] G. S. Colin A. Smith, Grace O. Maille, Elizabeth J. Want, Chuan Qin, Sunia A. Trauger, Theodore R. Brandon, Darlene E. Custodio, Ruben Abagyan, “METLIN: a Metabolite Mass Spectral Database,” *Ther. Drug Monit.*, vol. 27, no. 6, pp. 747–751, 2005.
- [6] S. G. Minoru Kanehisa, “KEGG : Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, 2000.
- [7] C. B. Joanna Godzien, Alberto Gil de la Fuente, Abraham Otero, “Metabolite Annotation and Identification,” in *Comprehensive Analytical Chemistry*, 1st ed., vol. 82, J. Jaumot, C. Bedia, and R. Tauler, Eds. Elsevier, 2018, pp. 415–445.
- [8] M. R. V. Warwick B. Dunn, Alexander Erban, Ralf J. M. Weber, Darren J. Creek, Marie Brown, Rainer Breitling, Thomas Hankemeier, Royston Goodacre, Steffen Neumann, Joachim Kopka, “Mass appeal: metabolite identification in mass

spectrometry-focused untargeted metabolomics,” *Metabolomics*, vol. 9, pp. 44–66, 2012.

[9] A. J. Arnald Alonso, Sara Marsal, “Analytical methods in untargeted metabolomics : state of the art in 2015,” *Front. Bioeng. Biotechnol.*, vol. 3, no. 25, 2015.

[10] H. W. R. Tsung-heng Tsai, Minkun Wang, “Preprocessing and Analysis of LC-MS-Based Proteomic Data,” *Methods Mol. Biol.*, vol. 1362, pp. 63–76, 2016.

[11] G. S. Colin A Smith, Elizabeth J Want, Grace O Maille, Ruben Abagyan, “XCMS : Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment , Matching , and Identification,” *Anal. Chem.*, vol. 78, no. 3, pp. 779–787, 2006.

[12] M. W. Röst, Hannes L. Röst, Timo Sachsenberg, Stephan Aiche, Chris Bielow, Hendrik Weisser, Fabian Aicheler, Sandro Andreotti, Hans-christian Ehrlich, Petra Gutenbrunner, Erhan Kenar, Xiao Liang, Sven Nahnsen, Lars Nilse, Julianus Pfeuffer, George Rosenberger, , D. Wojnar, W. E. Wolski, O. Schilling, and J. S. Choudhary, “OpenMS : a flexible open-source software platform for mass spectrometry data analysis,” *Nat. Methods*, vol. 13, no. 9, pp. 741–748, 2016.

[13] M. O. Tomá Pluskal, Sandra Castillo, Alejandro Villar-briones, “MZmine 2 : Modular framework for processing , visualizing , and analyzing mass spectrometry- based molecular profile data,” *BMC Bioinformatics*, vol. 11, no. 1, p. 395, 2010.

[14] C. B. Joanna Godzien, Michal Ciborowski, Santiago Angulo, “From numbers to a biological sense: How the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LC-MS,” *Electrophoresis*, vol. 34, no. 19, 2013.

[15] E. H. Johan Trygg, “Chemometrics in Metabonomics,” *J. Proteome Res.*, vol. 6, no. 2, pp. 469–479, 2007.

[16] M. R. V. Lloyd W Sumner, Alexander Amberg, Dave Barrett, Michael H Beale, Richard Beger, Clare A Daykin, Teresa W-M Fan, Oliver Fiehn, Royston Goodacre, Julian L Griffin, Thomas Hankemeier, Nigel Hardy, James Harnly, Richard Higashi, Joachim Kopka, Andrew N Lane, “Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative

(MSI),” *Metabolomics*, vol. 3, no. 3, pp. 211–221, 2007.

[17] A. O. Alberto Gil de la Fuente, Joanna Godzien, Sergio Saugar, Rodrigo Garcia-Carmona, Hasan Badran, David S. Wishart, Coral Barbas, “CEU Mass Mediator 3.0: A Metabolite Annotation Tool,” *Proteome Res.*, vol. 18, no. 2, pp. 797–802, 2018.

[18] H. W. R. Bin Zhou, Jinlian Wang, “MetaboSearch: Tool for Mass-Based Metabolite Identification Using Multiple Databases,” *PLoS One*, vol. 7, no. 6, 2012.

[19] S. Kim *et al.*, “PubChem in 2021: new data content and improved web interfaces,” vol. 49, no. November 2020, pp. 1388–1395, 2021.

[20] David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam Maciejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cum, “DrugBank 5.0: a major update to the DrugBank database for 2018,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. 1074–1082, 2018.

[21] R. A. Dominik Kopczynski, Nils Ho, Bing Peng, “Goslin: A Grammar of Succinct Lipid Nomenclature,” *Anal. Chem.*, vol. 92, no. 16, pp. 10957–10960, 2020.

[22] M. Sud *et al.*, “LMSD: LIPID MAPS structure database,” vol. 35, no. November 2006, pp. 527–532, 2007.

[23] L. Aimo *et al.*, “The SwissLipids knowledgebase for lipid biology,” vol. 31, no. May, pp. 2860–2866, 2015.

[24] “Python.” [Online]. Available: <https://www.python.org/>.

[25] “12. Virtual Environments and Packages.” [Online]. Available: <https://docs.python.org/3/tutorial/venv.html>.

[26] Z. Wang *et al.*, “Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease,” *Nature*, vol. 472, no. 7341, pp. 57–65, 2011.

[27] X. S. Li *et al.*, “Gut microbiota-dependent trimethylamine N-oxide in acute coronary syndromes: A prognostic marker for incident cardiovascular events beyond traditional risk factors,” *Eur. Heart J.*, vol. 38, no. 11, pp. 814–824, 2017.

Apéndice A

Tablas Procesadas con las Anotaciones Putativas

A continuación, se muestran las diferentes tablas con las posibles identificaciones que fueron procesadas en el experimento descrito en el Capítulo 3. Por razones de espacio, no se incluyó todo el contenido de las tablas. Se puede acceder a la información completa en el repositorio en GitHub del proyecto (<https://github.com/CNIC-Proteomics/TurboPutative/tree/master/test>).

A.1. Tabla Generada por Tagger

Experimental mass	Identifier	Adduct	mz Error (ppm)	Molecular Weight	Name	Peptide	Halogenated	Microbial	Drug	Food	Formula
336,8786	0	M+H	0	0	No compounds found for experimental mass 336.87859 and adduct: M+H						
336,8786	0	M+Na	0	0	No compounds found for experimental mass 336.87859 and adduct: M+Na						
336,8786	0	M+K	0	0	No compounds found for experimental mass 336.87859 and adduct: M+K						
336,8786	152122	M+H-H ₂ O	7	353,8843	Bithionol		x				C ₁₂ H ₆ Cl ₄ O ₂ S
336,8786	157817	M+H-H ₂ O	7	353,8843	Tetradifon		x				C ₁₂ H ₆ Cl ₄ O ₂ S
336,8786	0	M+H+HCOONa	0	0	No compounds found for experimental mass 336.87859 and adduct: M+H+HCOONa						
98,0577	0	M+H	0	0	No compounds found for experimental mass 98.05768 and adduct: M+H						
98,0577	107563	M+Na	1	75,0684	Trimethylamine-N-oxide			MC			C ₃ H ₉ NO
98,0577	117354	M+Na	1	75,0684	3-aminopropan-1-ol						C ₃ H ₉ NO
98,0577	58806	M+Na	1	75,0684	1-Amino-propan-2-ol						C ₃ H ₉ NO
98,0577	70315	M+Na	1	75,0684	2-(methylamino)ethan-1-ol						C ₃ H ₉ NO
98,0577	165739	M+Na	1	75,0684	(S)-(+)-2-Amino-1-propanol						C ₃ H ₉ NO
98,0577	165855	M+Na	1	75,0684	3-Aminopropanol						C ₃ H ₉ NO
98,0577	165856	M+Na	1	75,0684	2-Aminopropanol						C ₃ H ₉ NO

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

98,0577	159284	M+Na	1	75,0684	(R)-1-Aminopropan-2-ol; (R)-1-Amino-2-propanol						C3H9NO
98,0577	0	M+K	0	0	No compounds found for experimental mass 98.05768 and adduct: M+K						
98,0577	0	M+H-H2O	0	0	No compounds found for experimental mass 98.05768 and adduct: M+H-H2O						
98,0577	0	M+H+HCOONa	0	0	No compounds found for experimental mass 98.05768 and adduct: M+H+HCOONa						
189,0468	153648	M+H	1	188,0393	2-Chlorobiphenyl; 2-Monochlorobiphenyl; PCB 1		x				C12H9Cl
189,0468	159119	M+H	1	188,0393	4-Chlorobiphenyl; 1-Chloro-4-phenyl benzene; 4-Monochloro-biphenyl		x				C12H9Cl
189,0468	99417	M+Na	7	166,0565	Ethionamide				Drug		C8H10N2S
189,0468	0	M+K	0	0	No compounds found for experimental mass 189.0468 and adduct: M+K						
189,0468	158378	M+H-H2O	7	206,0514	Indolylmethylthiohydroximate						C10H10N2OS
189,0468	79879	M+H-H2O	7	206,0514	2-[(2-Furanylmethyl)thio]-6- methylpyrazine					Food	C10H10N2OS
189,0468	58247	M+H-H2O	7	206,0514	2-Methyl-3 or 5 or 6- (furfurylthio)pyrazine (mixture of isomers)						C10H10N2OS
189,0468	0	M+H+HCOONa	0	0	No compounds found for experimental mass 189.0468 and adduct: M+H+HCOONa						
205,0244	166585	M+H	9	204,0189	Honaucin A		x				C8H9ClO4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

205,0244	158225	M+Na	9	182,0368	Acenaphthenequinone; Acenaphthoquinone; 1,2-Diketoacenaphthene						C12H6O2
205,0244	0	M+K	0	0	No compounds found for experimental mass 205.02436 and adduct: M+K						
205,0244	0	M+H-H2O	0	0	No compounds found for experimental mass 205.02436 and adduct: M+H-H2O						
205,0244	27369	M+H+HCOONa	5	136,0291	5-chloropentanoic acid		x				C5H9O2Cl
80,5322	0	M+H	0	0	No compounds found for experimental mass 80.53221 and adduct: M+H						
80,5322	0	M+Na	0	0	No compounds found for experimental mass 80.53221 and adduct: M+Na						
80,5322	0	M+K	0	0	No compounds found for experimental mass 80.53221 and adduct: M+K						
80,5322	0	M+H-H2O	0	0	No compounds found for experimental mass 80.53221 and adduct: M+H-H2O						
80,5322	0	M+H+HCOONa	0	0	No compounds found for experimental mass 80.53221 and adduct: M+H+HCOONa						
117,055	27031	M+H	3	116,0473	3-keto valeric acid						C5H8O3
117,055	27148	M+H	3	116,0473	5-oxo-pentanoic acid						C5H8O3
117,055	27173	M+H	3	116,0473	2methyl-3-ketovaleric acid						C5H8O3
117,055	27030	M+H	3	116,0473	2-keto valeric acid						C5H8O3
117,055	27032	M+H	3	116,0473	Levulinic acid						C5H8O3
117,055	69545	M+H	3	116,0473	Acetoxyacetone					Food	C5H8O3

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

117,055	79290	M+H	3	116,0473	Ethyl pyruvate						C5H8O3
117,055	25548	M+H	3	116,0473	3-methyl-2-oxo-butanoic acid						C5H8O3
117,055	67562	M+H	3	116,0473	Methylacetoacetic acid						C5H8O3
117,055	83981	M+H	3	116,0473	(2E)-4-hydroxy-3-methylbut-2-enoic acid						C5H8O3
117,055	119091	M+H	3	116,0473	dimethyloxirane-2-carboxylic acid						C5H8O3
117,055	126887	M+H	3	116,0473	4-hydroxy-3-methylbut-2-enoic acid						C5H8O3
117,055	56251	M+H	3	116,0473	4-hydroxy-2-methylbut-2-enoic acid						C5H8O3
117,055	61151	M+H	3	116,0473	2,3-dimethyloxirane-2-carboxylic acid						C5H8O3
117,055	104679	M+H	3	116,0473	(2Z)-2-(hydroxymethyl)but-2-enoic acid						C5H8O3
117,055	107500	M+H	3	116,0473	2-(hydroxymethyl)but-2-enoic acid						C5H8O3
117,055	32034	M+H	3	116,0473	methyl 3-hydroxybut-2-enoate						C5H8O3
117,055	165033	M+H	3	116,0473	3-keto-2-methylbutyrate						C5H8O3
117,055	152401	M+Na	1	94,0657	N-Methylpyridinium						C6H8N
117,055	0	M+K	0	0	No compounds found for experimental mass 117.05497 and adduct: M+K						
117,055	151946	M+H-H2O	2	134,0579	Deoxyribose; 2-Deoxy-D-erythro-pentose; Thymine; 2-Deoxy-D-ribose						C5H10O4
117,055	156905	M+H-H2O	2	134,0579	1-Deoxy-D-xylulose						C5H10O4
117,055	26905	M+H-H2O	2	134,0579	2,3-dihydroxy-valeric acid						C5H10O4
117,055	26957	M+H-H2O	2	134,0579	(R)-2,3-Dihydroxy-isovalerate						C5H10O4
117,055	26969	M+H-H2O	2	134,0579	2,3-Dihydroxy-2-methylbutanoic acid						C5H10O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

117,055	73591	M+H-H ₂ O	2	134,0579	(R)-glycerol 1-acetate						C ₅ H ₁₀ O ₄
117,055	127641	M+H-H ₂ O	2	134,0579	1-Deoxy-D-xylulose						C ₅ H ₁₀ O ₄
117,055	60580	M+H-H ₂ O	2	134,0579	Deoxyribose						C ₅ H ₁₀ O ₄
117,055	134513	M+H-H ₂ O	2	134,0579	3-hydroxy-2-(hydroxymethyl)-2-methylpropanoic acid						C ₅ H ₁₀ O ₄
117,055	150571	M+H-H ₂ O	2	134,0579	2-Deoxy-alpha-D-ribose						C ₅ H ₁₀ O ₄
117,055	157489	M+H-H ₂ O	2	134,0579	2-Deoxy-L-arabinose						C ₅ H ₁₀ O ₄
117,055	161603	M+H-H ₂ O	2	134,0579	2,3-Dihydroxy-3-methylbutanoate; 2,3-Dihydroxy-isovalerate; 2,3-Dihydroxy-isovaleric acid						C ₅ H ₁₀ O ₄
117,055	165059	M+H-H ₂ O	2	134,0579	2-Deoxy-D-Ribose						C ₅ H ₁₀ O ₄
117,055	165060	M+H-H ₂ O	2	134,0579	1-Deoxy-D-xylulose						C ₅ H ₁₀ O ₄
117,055	165061	M+H-H ₂ O	2	134,0579	2,3-Dihydroxyvaleric acid						C ₅ H ₁₀ O ₄
117,055	165062	M+H-H ₂ O	2	134,0579	2,3-Dihydroxy-3-methylbutyric acid						C ₅ H ₁₀ O ₄
117,055	44842	M+H-H ₂ O	8	134,0592	Dihydropteridine						C ₆ H ₆ N ₄
117,055	0	M+H+HCOONa	0	0	No compounds found for experimental mass 117.05497 and adduct: M+H+HCOONa						
95,0606	152438	M+H	2	94,0531	4-Aminopyridine; 4-AP						C ₅ H ₆ N ₂
95,0606	120308	M+H	2	94,0531	Methylpyrazine					Food	C ₅ H ₆ N ₂
95,0606	0	M+Na	0	0	No compounds found for experimental mass 95.06057 and adduct: M+Na						
95,0606	0	M+K	0	0	No compounds found for experimental mass 95.06057 and adduct: M+K						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

95,0606	0	M+H-H ₂ O	0	0	No compounds found for experimental mass 95.06057 and adduct: M+H-H ₂ O						
95,0606	0	M+H+HCOONa	0	0	No compounds found for experimental mass 95.06057 and adduct: M+H+HCOONa						
162,0765	157709	M+H	3	161,0688	D-Alanyl-(R)-lactate; D-Alanyl-D-lactate; (2R)-2-(D-Alanyloxy)propanoic acid; (R)-2-[(R)-2-Aminopropanoyloxy]propanoic acid						C ₆ H ₁₁ NO ₄
162,0765	154712	M+H	3	161,0688	3-Amino-2,3-dideoxy-scylo- inosose; 2-Deoxy-3-keto-scylo-inosamine			MC			C ₆ H ₁₁ NO ₄
162,0765	157301	M+H	3	161,0688	4-Methyl-L-glutamate						C ₆ H ₁₁ NO ₄
162,0765	150972	M+H	3	161,0688	beta-Alanopine; N-(D-1-Carboxyethyl)-beta-alanine						C ₆ H ₁₁ NO ₄
162,0765	151507	M+H	3	161,0688	L-Glutamate methylester						C ₆ H ₁₁ NO ₄
162,0765	165098	M+H	3	161,0688	Acetylhomoserine						C ₆ H ₁₁ NO ₄
162,0765	79135	M+H	3	161,0688	N-Acetylthreonine						C ₆ H ₁₁ NO ₄
162,0765	145448	M+H	3	161,0688	(±)-2,2'-Iminobispropanoic acid					Food	C ₆ H ₁₁ NO ₄
162,0765	126061	M+H	3	161,0688	Acetylhomoserine						C ₆ H ₁₁ NO ₄
162,0765	128184	M+H	3	161,0688	Glutamate, gamma-methyl ester						C ₆ H ₁₁ NO ₄
162,0765	116155	M+H	3	161,0688	N-methyl-L-glutamic Acid						C ₆ H ₁₁ NO ₄
162,0765	27622	M+H	3	161,0688	2-Aminoadipic acid						C ₆ H ₁₁ NO ₄
162,0765	105769	M+H	3	161,0688	hydroxybutyrylglycine						C ₆ H ₁₁ NO ₄
162,0765	180841	M+H	3	161,0688	L-2-aminoadipic acid						C ₆ H ₁₁ NO ₄
162,0765	162588	M+H	3	161,0688	O-Acetyl-L-homoserine; O-Acetylhomoserine						C ₆ H ₁₁ NO ₄

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

162,0765	0	M+Na	0	0	No compounds found for experimental mass 162.07649 and adduct: M+Na						
162,0765	0	M+K	0	0	No compounds found for experimental mass 162.07649 and adduct: M+K						
162,0765	156462	M+H-H ₂ O	2	179,0794	D-Galactosamine; D-Chondrosamine; 2-Amino-2-deoxy-D-galactose						C ₆ H ₁₃ NO ₅
162,0765	163149	M+H-H ₂ O	2	179,0794	Nojirimycin; 5-Amino-5-deoxy-D-glucopyranose			MC			C ₆ H ₁₃ NO ₅
162,0765	162443	M+H-H ₂ O	2	179,0794	Kanosamine; D-Kanosamine; 3-Amino-3-deoxy-D-glucose			MC			C ₆ H ₁₃ NO ₅
162,0765	161935	M+H-H ₂ O	2	179,0794	1-Amino-1-deoxy-scylo-inositol; scylo-Inosamine			MC			C ₆ H ₁₃ NO ₅
162,0765	122199	M+H-H ₂ O	2	179,0794	beta-D-Glucosamine						C ₆ H ₁₃ NO ₅
162,0765	117693	M+H-H ₂ O	2	179,0794	Fructosamine						C ₆ H ₁₃ NO ₅
162,0765	121533	M+H-H ₂ O	2	179,0794	D-Glucosamine						C ₆ H ₁₃ NO ₅
162,0765	158238	M+H-H ₂ O	2	179,0794	D-Mannosamine; 2-Amino-2-deoxy-D-mannose						C ₆ H ₁₃ NO ₅
162,0765	150567	M+H-H ₂ O	2	179,0794	D-Talosamine; 2-Amino-2-deoxy-D-talopyranose						C ₆ H ₁₃ NO ₅
162,0765	158262	M+H-H ₂ O	2	179,0794	neo-Inosamine-2; 2L-2-Amino-2-deoxy-neo-inositol			MC			C ₆ H ₁₃ NO ₅
162,0765	152384	M+H-H ₂ O	2	179,0794	L-Altrosamine; 2-Amino-2-deoxy-L-altropyranose						C ₆ H ₁₃ NO ₅
162,0765	150900	M+H-H ₂ O	2	179,0794	D-Gulosamine; 2-Amino-2-deoxy-D-gulopyranose						C ₆ H ₁₃ NO ₅
162,0765	150993	M+H-H ₂ O	2	179,0794	L-Idosamine; 2-Amino-2-deoxy-L-idopyranose						C ₆ H ₁₃ NO ₅

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

162,0765	155900	M+H-H ₂ O	2	179,0794	D-Allosamine; 2-Amino-2-deoxy-D-allopyranose						C ₆ H ₁₃ NO ₅
162,0765	120004	M+H-H ₂ O	6	179,0807	7-Aminomethyl-7-carbaguanine						C ₇ H ₉ N ₅ O
162,0765	0	M+H+HCOONa	0	0	No compounds found for experimental mass 162.07649 and adduct: M+H+HCOONa						
95,001	0	M+H	0	0	No compounds found for experimental mass 95.00101 and adduct: M+H						
95,001	0	M+Na	0	0	No compounds found for experimental mass 95.00101 and adduct: M+Na						
95,001	0	M+K	0	0	No compounds found for experimental mass 95.00101 and adduct: M+K						
95,001	0	M+H-H ₂ O	0	0	No compounds found for experimental mass 95.00101 and adduct: M+H-H ₂ O						
95,001	0	M+H+HCOONa	0	0	No compounds found for experimental mass 95.00101 and adduct: M+H+HCOONa						
498,2905	0	M+H	0	0	No compounds found for experimental mass 498.29047 and adduct: M+H						
498,2905	179482	M+Na	7	475,3046	(S)-MG132						C ₂₆ H ₄₁ N ₃ O ₅
498,2905	179483	M+Na	7	475,3046	(R)-MG132						C ₂₆ H ₄₁ N ₃ O ₅
498,2905	153542	M+Na	1	475,3006	Netilmicin				Drug		C ₂₁ H ₄₁ N ₅ O ₇
498,2905	76624	M+Na	1	475,3006	Netilmicin				Drug		C ₂₁ H ₄₁ N ₅ O ₇
498,2905	0	M+K	0	0	No compounds found for experimental mass 498.29047 and adduct: M+K						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

498,2905	164530	M+H-H ₂ O	4	515,2917	Taurocholic acid						C26H45NO7S
498,2905	102419	M+H-H ₂ O	4	515,2917	Taurohyocholate						C26H45NO7S
498,2905	83105	M+H-H ₂ O	4	515,2917	Tauro-b-muricholic acid						C26H45NO7S
498,2905	41189	M+H-H ₂ O	4	515,2917	N-Taurocholate						C26H45NO7S
498,2905	41204	M+H-H ₂ O	4	515,2917	Taurallocholic acid						C26H45NO7S
498,2905	57855	M+H-H ₂ O	4	515,2917	Tauroursocholic acid						C26H45NO7S
498,2905	41196	M+H-H ₂ O	4	515,2917	Taurohyocholic acid						C26H45NO7S
498,2905	41197	M+H-H ₂ O	4	515,2917	Tauro-b-muricholic acid						C26H45NO7S
498,2905	41199	M+H-H ₂ O	4	515,2917	Tauroursocholic acid						C26H45NO7S
498,2905	41201	M+H-H ₂ O	4	515,2917	Taurallocholic acid						C26H45NO7S
498,2905	162382	M+H-H ₂ O	4	515,2917	Taurocholate; Taurocholic acid; Cholytaurine						C26H45NO7S
498,2905	0	M+H+HCOONa	0	0	No compounds found for experimental mass 498.29047 and adduct: M+H+HCOONa						
244,6389	0	M+H	0	0	No compounds found for experimental mass 244.6389 and adduct: M+H						
244,6389	0	M+Na	0	0	No compounds found for experimental mass 244.6389 and adduct: M+Na						
244,6389	0	M+K	0	0	No compounds found for experimental mass 244.6389 and adduct: M+K						
244,6389	0	M+H-H ₂ O	0	0	No compounds found for experimental mass 244.6389 and adduct: M+H-H ₂ O						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

244,6389	0	M+H+HCOONa	0	0	No compounds found for experimental mass 244.6389 and adduct: M+H+HCOONa						
357,28	25858	M+H	3	356,2715	4,8,12,15,19,21-tetracosahexaenoic acid						C24H36O2
357,28	26388	M+H	3	356,2715	tetracosahexaenoic acid						C24H36O2
357,28	40744	M+H	3	356,2715	5beta-Chola-3,11-dien-24-oic Acid						C24H36O2
357,28	40795	M+H	3	356,2715	chola-4,6-dien-24-oic acid						C24H36O2
357,28	40351	M+H	3	356,2715	(1S)-1-hydroxy-23,24-didehydro-25,26,27-trinorcalciol						C24H36O2
357,28	40685	M+H	3	356,2715	5beta-Chola-8(14),11-dien-24-oic Acid						C24H36O2
357,28	40687	M+H	3	356,2715	5beta-Chola-7,9(11)-dien-24-oic Acid						C24H36O2
357,28	26406	M+H	3	356,2715	THA				Drug		C24H36O2
357,28	97730	M+H	3	356,2715	Tetracosahexaenoic acid, n-3						C24H36O2
357,28	180800	M+H	3	356,2715	C24:6						C24H36O2
357,28	155668	M+H	3	356,2715	(4Z,7Z,10Z,13Z,16Z,19Z)-Docosahexaenoic acid ethyl ester; DHA ethyl ester						C24H36O2
357,28	172421	M+H	3	356,2715	4,8,12,15,19,21-Tetracosahexaenoic acid						C24H36O2
357,28	172422	M+H	3	356,2715	5β-Chola-8(14),11-dien-24-oic Acid						C24H36O2
357,28	172423	M+H	3	356,2715	5β-Chola-7,9(11)-dien-24-oic Acid						C24H36O2
357,28	172424	M+H	3	356,2715	5β-Chola-3,11-dien-24-oic Acid						C24H36O2
357,28	26505	M+H	3	356,2715	(6E,9E,12E,15E,18E,21E)-tetracosahexaenoic acid						C24H36O2
357,28	172425	M+H	3	356,2715	Docosahexaenoic Acid ethyl ester						C24H36O2
357,28	164347	M+H	3	356,2715	Tetracosahexaenoic acid						C24H36O2

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

357,28	0	M+Na	0	0	No compounds found for experimental mass 357.28002 and adduct: M+Na						
357,28	0	M+K	0	0	No compounds found for experimental mass 357.28002 and adduct: M+K						
357,28	40564	M+H-H ₂ O	3	374,2821	3beta-Hydroxy-chol-5-en-24-oic Acid						C24H38O3
357,28	60039	M+H-H ₂ O	3	374,2821	2-(10-Heptadecenyl)-6-hydroxybenzoic acid					Food	C24H38O3
357,28	65725	M+H-H ₂ O	3	374,2821	D8'-Merulinic acid C					Food	C24H38O3
357,28	17679	M+H-H ₂ O	3	374,2821	6-((Z)-pentadec-8-en-1-yl)salicylic acid						C24H38O3
357,28	39697	M+H-H ₂ O	3	374,2821	1alpha,24-dihydroxy-25,26,27-trinorvitamin D3 / 1alpha,24-dihydroxy-25,26,27-trinorcholecalciferol						C24H38O3
357,28	40491	M+H-H ₂ O	3	374,2821	3-Oxo-5beta-cholan-24-oic Acid						C24H38O3
357,28	40492	M+H-H ₂ O	3	374,2821	3-Oxo-5alpha-cholan-24-oic Acid						C24H38O3
357,28	40493	M+H-H ₂ O	3	374,2821	6-Oxo-5beta-cholan-24-oic Acid						C24H38O3
357,28	40494	M+H-H ₂ O	3	374,2821	6-Oxo-5alpha-cholan-24-oic Acid						C24H38O3
357,28	40495	M+H-H ₂ O	3	374,2821	7-Oxo-5beta-cholan-24-oic Acid						C24H38O3
357,28	40496	M+H-H ₂ O	3	374,2821	12-Oxo-5beta-cholan-24-oic Acid						C24H38O3
357,28	40497	M+H-H ₂ O	3	374,2821	12-Oxo-5alpha-cholan-24-oic Acid						C24H38O3
357,28	26945	M+H-H ₂ O	3	374,2821	15-hydroxy-tetracos-6,9,12,16,18-pentaenoic acid						C24H38O3
357,28	40561	M+H-H ₂ O	3	374,2821	3alpha-Hydroxychol-4-en-24-oic Acid						C24H38O3
357,28	40562	M+H-H ₂ O	3	374,2821	3beta-Hydroxychol-4-en-24-oic Acid						C24H38O3

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

357,28	40563	M+H-H ₂ O	3	374,2821	3alpha-Hydroxychol-5-en-24-oic Acid						C24H38O3
357,28	40565	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5beta-chol-6-en-24-oic Acid						C24H38O3
357,28	40566	M+H-H ₂ O	3	374,2821	3beta-Hydroxy-5beta-chol-6-en-24-oic Acid						C24H38O3
357,28	40567	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5beta-chol-7-en-24-oic Acid						C24H38O3
357,28	40568	M+H-H ₂ O	3	374,2821	3beta-Hydroxy-5alpha-chol-7-en-24-oic Acid						C24H38O3
357,28	40569	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5beta-chol-8-en-24-oic Acid						C24H38O3
357,28	40570	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5beta-chol-8(14)-en-24-oic Acid						C24H38O3
357,28	40571	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5beta-chol-9(11)-en-24-oic Acid						C24H38O3
357,28	40572	M+H-H ₂ O	3	374,2821	3beta-Hydroxy-5beta-chol-9(11)-en-24-oic Acid						C24H38O3
357,28	40573	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5beta-chol-11-en-24-oic Acid						C24H38O3
357,28	40574	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5beta-chol-14-en-24-oic Acid						C24H38O3
357,28	40575	M+H-H ₂ O	3	374,2821	7alpha-Hydroxy-5beta-chol-3-en-24-oic Acid						C24H38O3
357,28	40576	M+H-H ₂ O	3	374,2821	12alpha-Hydroxy-5beta-chol-3-en-24-oic Acid						C24H38O3
357,28	40632	M+H-H ₂ O	3	374,2821	7-Oxo-5alpha-cholan-24-oic Acid						C24H38O3
357,28	40640	M+H-H ₂ O	3	374,2821	7alpha-Hydroxy-5beta-chol-2-en-24-oic Acid						C24H38O3
357,28	40648	M+H-H ₂ O	3	374,2821	3beta-Hydroxy-5beta-chol-1-en-24-oic Acid						C24H38O3

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

357,28	40651	M+H-H ₂ O	3	374,2821	11-Oxo-5beta-cholan-24-oic Acid						C24H38O3
357,28	40678	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5beta,14beta-chol-8-en-24-oic Acid						C24H38O3
357,28	40681	M+H-H ₂ O	3	374,2821	(20S)-3beta-Hydroxychol-5-en-24-oic Acid						C24H38O3
357,28	40683	M+H-H ₂ O	3	374,2821	(20S,22E)-3beta-Hydroxy-5alpha-chol-22-en-24-oic Acid						C24H38O3
357,28	40693	M+H-H ₂ O	3	374,2821	3alpha-Hydroxy-5alpha-chol-6-en-24-oic Acid						C24H38O3
357,28	40703	M+H-H ₂ O	3	374,2821	15-Oxo-5beta-cholan-24-oic Acid						C24H38O3
357,28	151054	M+H-H ₂ O	3	374,2821	3beta-[(Tetrahydro-2H-pyran-2-yl)oxy]androst-5-en-17beta-ol; Androst-5-ene-3beta,17beta-diol 3-tetrahydropyranyl ether						C24H38O3
357,28	153649	M+H-H ₂ O	3	374,2821	3beta,21-Dihydroxy-4,4,14-trimethyl-5alpha-pregn-8-en-20-one						C24H38O3
357,28	173988	M+H-H ₂ O	3	374,2821	3-Oxo-5β-cholan-24-oic Acid						C24H38O3
357,28	173989	M+H-H ₂ O	3	374,2821	6-Oxo-5β-cholan-24-oic Acid						C24H38O3
357,28	173990	M+H-H ₂ O	3	374,2821	6-Oxo-5a-cholan-24-oic Acid						C24H38O3
357,28	173991	M+H-H ₂ O	3	374,2821	7-Oxo-5β-cholan-24-oic Acid						C24H38O3
357,28	173992	M+H-H ₂ O	3	374,2821	12-Oxo-5β-cholan-24-oic Acid						C24H38O3
357,28	173993	M+H-H ₂ O	3	374,2821	12-Oxo-5a-cholan-24-oic Acid						C24H38O3
357,28	173994	M+H-H ₂ O	3	374,2821	3a-Hydroxychol-4-en-24-oic Acid						C24H38O3
357,28	173995	M+H-H ₂ O	3	374,2821	3β-Hydroxychol-4-en-24-oic Acid						C24H38O3
357,28	173996	M+H-H ₂ O	3	374,2821	3a-Hydroxychol-5-en-24-oic Acid						C24H38O3
357,28	173997	M+H-H ₂ O	3	374,2821	3β-Hydroxy-chol-5-en-24-oic Acid						C24H38O3
357,28	173998	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5β-chol-6-en-24-oic Acid						C24H38O3

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

357,28	173999	M+H-H ₂ O	3	374,2821	3 β -Hydroxy-5 β -chol-6-en-24-oic Acid						C24H38O3
357,28	174000	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5 β -chol-7-en-24-oic Acid						C24H38O3
357,28	174001	M+H-H ₂ O	3	374,2821	3 β -Hydroxy-5a-chol-7-en-24-oic Acid						C24H38O3
357,28	174002	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5 β -chol-8-en-24-oic Acid						C24H38O3
357,28	174003	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5 β -chol-8(14)-en-24-oic Acid						C24H38O3
357,28	174004	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5 β -chol-9(11)-en-24-oic Acid						C24H38O3
357,28	174005	M+H-H ₂ O	3	374,2821	3 β -Hydroxy-5 β -chol-9(11)-en-24-oic Acid						C24H38O3
357,28	174006	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5 β -chol-11-en-24-oic Acid						C24H38O3
357,28	174007	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5 β -chol-14-en-24-oic Acid						C24H38O3
357,28	174008	M+H-H ₂ O	3	374,2821	7a-Hydroxy-5 β -chol-3-en-24-oic Acid						C24H38O3
357,28	174009	M+H-H ₂ O	3	374,2821	12a-Hydroxy-5 β -chol-3-en-24-oic Acid						C24H38O3
357,28	174010	M+H-H ₂ O	3	374,2821	7-Oxo-5a-cholan-24-oic Acid						C24H38O3
357,28	174011	M+H-H ₂ O	3	374,2821	7a-Hydroxy-5 β -chol-2-en-24-oic Acid						C24H38O3
357,28	174012	M+H-H ₂ O	3	374,2821	3 β -Hydroxy-5 β -chol-1-en-24-oic Acid						C24H38O3
357,28	174013	M+H-H ₂ O	3	374,2821	11-Oxo-5 β -cholan-24-oic Acid						C24H38O3
357,28	174014	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5 β ,14 β -chol-8-en-24-oic Acid						C24H38O3

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

357,28	174015	M+H-H ₂ O	3	374,2821	(20S)-3 β -Hydroxychol-5-en-24-oic Acid						C24H38O3
357,28	174016	M+H-H ₂ O	3	374,2821	(20S,22E)-3 β -Hydroxy-5a-chol-22-en-24-oic Acid						C24H38O3
357,28	174017	M+H-H ₂ O	3	374,2821	3a-Hydroxy-5a-chol-6-en-24-oic Acid						C24H38O3
357,28	174018	M+H-H ₂ O	3	374,2821	15-Oxo-5 β -cholan-24-oic Acid						C24H38O3
357,28	174019	M+H-H ₂ O	6	374,2855	2-thio-Acetyl MAGE						C21H42O3S
357,28	0	M+H+HCOONa	0	0	No compounds found for experimental mass 357.28002 and adduct: M+H+HCOONa						
410,3279	176882	M+H	3	409,3192	PGF2a diethyl amide						C24H43NO4
410,3279	176883	M+H	3	409,3192	Lumula						C24H43NO4
410,3279	0	M+Na	0	0	No compounds found for experimental mass 410.32791 and adduct: M+Na						
410,3279	0	M+K	0	0	No compounds found for experimental mass 410.32791 and adduct: M+K						
410,3279	0	M+H-H ₂ O	0	0	No compounds found for experimental mass 410.32791 and adduct: M+H-H ₂ O						
410,3279	0	M+H+HCOONa	0	0	No compounds found for experimental mass 410.32791 and adduct: M+H+HCOONa						
339,6053	0	M+H	0	0	No compounds found for experimental mass 339.60535 and adduct: M+H						
339,6053	0	M+Na	0	0	No compounds found for experimental mass 339.60535 and adduct: M+Na						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

339,6053	0	M+K	0	0	No compounds found for experimental mass 339.60535 and adduct: M+K						
339,6053	0	M+H-H ₂ O	0	0	No compounds found for experimental mass 339.60535 and adduct: M+H-H ₂ O						
339,6053	0	M+H+HCOONa	0	0	No compounds found for experimental mass 339.60535 and adduct: M+H+HCOONa						
367,1815	0	M+H	0	0	No compounds found for experimental mass 367.18154 and adduct: M+H						
367,1815	160460	M+Na	10	344,1889	(+)TAN-67						C23H24N2O
367,1815	152527	M+Na	10	344,1889	(-)TAN-67						C23H24N2O
367,1815	0	M+K	0	0	No compounds found for experimental mass 367.18154 and adduct: M+K						
367,1815	174785	M+H-H ₂ O	2	384,1838	JWH 200						C25H24N2O2
367,1815	174786	M+H-H ₂ O	2	384,1838	JWH 200 2'-naphthyl isomer						C25H24N2O2
367,1815	100474	M+H+HCOONa	3	298,1879	Diphthamide						C13H24N5O3
434,7922	0	M+H	0	0	No compounds found for experimental mass 434.79218 and adduct: M+H						
434,7922	0	M+Na	0	0	No compounds found for experimental mass 434.79218 and adduct: M+Na						
434,7922	0	M+K	0	0	No compounds found for experimental mass 434.79218 and adduct: M+K						
434,7922	0	M+H-H ₂ O	0	0	No compounds found for experimental mass 434.79218 and adduct: M+H-H ₂ O						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

434,7922	0	M+H+HCOONa	0	0	No compounds found for experimental mass 434.79218 and adduct: M+H+HCOONa						
387,3272	39714	M+H	4	386,3185	25-hydroxy-3-deoxy-2-oxavitamin D3 / 25-hydroxy-3-deoxy-2-oxacholecalciferol						C26H42O2
387,3272	26415	M+H	4	386,3185	Shibic acid						C26H42O2
387,3272	26429	M+H	4	386,3185	26:5(8Z,11Z,14Z,17Z,20Z)						C26H42O2
387,3272	109668	M+H	4	386,3185	1-Phenyl-1,3-eicosanedione					Food	C26H42O2
387,3272	180772	M+H	4	386,3185	C26:5						C26H42O2
387,3272	174926	M+H	4	386,3185	3-beta-entanthoxyandrost-1-en-17-one						C26H42O2
387,3272	0	M+Na	0	0	No compounds found for experimental mass 387.32718 and adduct: M+Na						
387,3272	0	M+K	0	0	No compounds found for experimental mass 387.32718 and adduct: M+K						
387,3272	40807	M+H-H2O	3	404,329	Dihomolithocholic acid						C26H44O3
387,3272	39748	M+H-H2O	3	404,329	1alpha,25-dihydroxy-19-norvitamin D3 / 1alpha,25-dihydroxy-19-norcholecalciferol						C26H44O3
387,3272	39749	M+H-H2O	3	404,329	1,25-dihydroxy-2-nor-1,2-secovitamin D3 / 1,25-dihydroxy-2-nor-1,2-secocholecalciferol						C26H44O3
387,3272	40342	M+H-H2O	3	404,329	2-Nor-1,3-seco-1alpha,25-dihydroxyvitamin D3						C26H44O3
387,3272	0	M+H+HCOONa	0	0	No compounds found for experimental mass 387.32718 and adduct: M+H+HCOONa						
142,0268	102283	M+H	2	141,0193	(2E,4E)-5-carboxypenta-2,4-dienoate						C6H5O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

142,0268	112640	M+H	3	141,0191	Ethanolamine phosphate						C2H8NO4P
142,0268	153913	M+H	3	141,0191	(2-Amino-1-hydroxyethyl)phosphonate; 1-Hydroxy-2-aminoethylphosphonate						C2H8NO4P
142,0268	127260	M+H	5	141,0188	5-Hydroxymethyl-2-furanoate						C6H5O4
142,0268	148609	M+Na	4	119,0371	1,2-Diacylglycerol-LD-PI-pool						C7H5NO
142,0268	151840	M+K	3	103,0633	n-Propyl carbamate						C4H9NO2
142,0268	165029	M+K	3	103,0633	2R-amino-butanoic acid						C4H9NO2
142,0268	153581	M+K	3	103,0633	HBA						C4H9NO2
142,0268	123399	M+K	3	103,0633	O-Acetyethanolamine					Food	C4H9NO2
142,0268	27400	M+K	3	103,0633	2-Aminobutyric acid						C4H9NO2
142,0268	27405	M+K	3	103,0633	4-Aminobutyric acid						C4H9NO2
142,0268	27409	M+K	3	103,0633	2R-amino-butanoic acid						C4H9NO2
142,0268	27414	M+K	3	103,0633	(R)-b-amino-isobutyric acid						C4H9NO2
142,0268	27415	M+K	3	103,0633	(S)-b-amino-isobutyric acid						C4H9NO2
142,0268	27418	M+K	3	103,0633	2-amino-isobutyric acid						C4H9NO2
142,0268	27419	M+K	3	103,0633	3-amino-isobutanoic acid						C4H9NO2
142,0268	43122	M+K	3	103,0633	N,N-Dimethylglycine						C4H9NO2
142,0268	142984	M+K	3	103,0633	N-Ethylglycine						C4H9NO2
142,0268	123815	M+K	3	103,0633	Butyl nitrite						C4H9NO2
142,0268	46261	M+K	3	103,0633	3-Aminobutyric acid						C4H9NO2
142,0268	80616	M+K	3	103,0633	N-Methylalanine						C4H9NO2
142,0268	0	M+H-H2O	0	0	No compounds found for experimental mass 142.02682 and adduct: M+H-H2O						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

142,0268	0	M+H+HCOONa	0	0	No compounds found for experimental mass 142.02682 and adduct: M+H+HCOONa						
311,2959	31760	M+H	5	310,2872	(E)-3,7-Dimethyl-6-octenyl 3,7-dimethyl-2,6-octadienoate						C20H38O2
311,2959	31533	M+H	5	310,2872	13E-Octadecenyl acetate						C20H38O2
311,2959	31534	M+H	5	310,2872	2E-Octadecenyl acetate						C20H38O2
311,2959	31538	M+H	5	310,2872	11Z-Octadecenyl acetate						C20H38O2
311,2959	31539	M+H	5	310,2872	13Z-Octadecenyl acetate						C20H38O2
311,2959	31540	M+H	5	310,2872	9Z-Octadecenyl acetate						C20H38O2
311,2959	26442	M+H	5	310,2872	20:1(7Z)						C20H38O2
311,2959	28235	M+H	5	310,2872	Prostanoic acid skeleton						C20H38O2
311,2959	25762	M+H	5	310,2872	17Z-octadecenoic acid						C20H38O2
311,2959	25766	M+H	5	310,2872	trans-gondoic acid						C20H38O2
311,2959	25767	M+H	5	310,2872	14Z-eicosenoic acid						C20H38O2
311,2959	26030	M+H	5	310,2872	13-eicosenoic acid						C20H38O2
311,2959	26032	M+H	5	310,2872	15-eicosenoic acid						C20H38O2
311,2959	26033	M+H	5	310,2872	15Z-eicosenoic acid						C20H38O2
311,2959	26034	M+H	5	310,2872	5E-eicosenoic acid						C20H38O2
311,2959	26035	M+H	5	310,2872	5Z-eicosenoic acid						C20H38O2
311,2959	26036	M+H	5	310,2872	8Z-eicosenoic acid						C20H38O2
311,2959	26301	M+H	5	310,2872	Gadelaidic acid						C20H38O2
311,2959	488	M+H	5	310,2872	2Z-Phytanoic Acid						C20H38O2
311,2959	32247	M+H	5	310,2872	20-Eicosanolide						C20H38O2
311,2959	506	M+H	5	310,2872	2E-Phytanoic acid						C20H38O2
311,2959	507	M+H	5	310,2872	2E-Phytenic acid						C20H38O2

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

311,2959	25764	M+H	5	310,2872	cis-Gadoleic acid						C20H38O2
311,2959	31548	M+H	5	310,2872	9-Octadecenyl acetate						C20H38O2
311,2959	130670	M+H	5	310,2872	4,6-Icosanedione					Food	C20H38O2
311,2959	117889	M+H	5	310,2872	5,7-Icosanedione					Food	C20H38O2
311,2959	25765	M+H	5	310,2872	cis-gondoic acid						C20H38O2
311,2959	141996	M+H	5	310,2872	6,8-Icosanedione					Food	C20H38O2
311,2959	26031	M+H	5	310,2872	Paullinic acid						C20H38O2
311,2959	31974	M+H	5	310,2872	Ethyl oleate						C20H38O2
311,2959	50424	M+H	5	310,2872	11-Eicosenoic acid						C20H38O2
311,2959	180636	M+H	5	310,2872	C20:1						C20H38O2
311,2959	26625	M+H	5	310,2872	6Z-Eicosenoic acid						C20H38O2
311,2959	26630	M+H	5	310,2872	3E-Eicosenoic acid						C20H38O2
311,2959	26645	M+H	5	310,2872	16Z-eicosenoic acid						C20H38O2
311,2959	169380	M+H	5	310,2872	Phytenic acid						C20H38O2
311,2959	169381	M+H	5	310,2872	17Z-Octadecenoic acid						C20H38O2
311,2959	169382	M+H	5	310,2872	14(Z)-Eicosenoic Acid						C20H38O2
311,2959	169383	M+H	5	310,2872	13-eicosenoic acid						C20H38O2
311,2959	169384	M+H	5	310,2872	13Z-eicosenoic acid						C20H38O2
311,2959	169385	M+H	5	310,2872	15-eicosenoic acid						C20H38O2
311,2959	169386	M+H	5	310,2872	15Z-eicosenoic acid						C20H38O2
311,2959	169387	M+H	5	310,2872	5E-eicosenoic acid						C20H38O2
311,2959	169388	M+H	5	310,2872	5Z-eicosenoic acid						C20H38O2
311,2959	169389	M+H	5	310,2872	8Z-eicosenoic acid						C20H38O2
311,2959	169390	M+H	5	310,2872	9-Eicosenoic acid						C20H38O2
311,2959	169391	M+H	5	310,2872	Oleic Acid ethyl ester						C20H38O2

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

311,2959	169392	M+H	5	310,2872	?2-cis Eicosenoic Acid						C20H38O2
311,2959	26620	M+H	5	310,2872	10Z-Eicosenoic acid						C20H38O2
311,2959	0	M+Na	0	0	No compounds found for experimental mass 311.29586 and adduct: M+Na						
311,2959	0	M+K	0	0	No compounds found for experimental mass 311.29586 and adduct: M+K						
311,2959	181444	M+H-H2O	4	328,2977	11-methoxy-nonadecanoic acid						C20H40O3
311,2959	26736	M+H-H2O	4	328,2977	3-hydroxy-eicosanoic acid						C20H40O3
311,2959	72196	M+H-H2O	4	328,2977	12-hydroxyicosanoic acid						C20H40O3
311,2959	91397	M+H-H2O	4	328,2977	15-hydroxyicosanoic acid						C20H40O3
311,2959	105247	M+H-H2O	4	328,2977	Polyoxyethylene 40 monostearate						C20H40O3
311,2959	26735	M+H-H2O	4	328,2977	2-hydroxy-eicosanoic acid						C20H40O3
311,2959	494	M+H-H2O	4	328,2977	2-hydroxyphytanic acid					Food	C20H40O3
311,2959	26737	M+H-H2O	4	328,2977	20-hydroxy-eicosanoic acid						C20H40O3
311,2959	26843	M+H-H2O	4	328,2977	3R-hydroxy-eicosanoic acid						C20H40O3
311,2959	498	M+H-H2O	4	328,2977	(2S)-2-hydroxyphytanic acid						C20H40O3
311,2959	170361	M+H-H2O	4	328,2978	3R-hydroxy-eicosanoic acid						C20H40O3
311,2959	0	M+H+HCOONa	0	0	No compounds found for experimental mass 311.29586 and adduct: M+H+HCOONa						
115,0546	0	M+H	0	0	No compounds found for experimental mass 115.05456 and adduct: M+H						
115,0546	0	M+Na	0	0	No compounds found for experimental mass 115.05456 and adduct: M+Na						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

115,0546	0	M+K	0	0	No compounds found for experimental mass 115.05456 and adduct: M+K						
115,0546	107276	M+H-H ₂ O	2	132,0575	Atropaldehyde						C ₉ H ₈ O
115,0546	99879	M+H-H ₂ O	2	132,0575	Indanone						C ₉ H ₈ O
115,0546	42600	M+H-H ₂ O	2	132,0575	2-Methylcumarone						C ₉ H ₈ O
115,0546	118383	M+H-H ₂ O	2	132,0575	(Z)-3-Phenyl-2-propenal						C ₉ H ₈ O
115,0546	126146	M+H-H ₂ O	2	132,0575	Cinnamaldehyde					Food	C ₉ H ₈ O
115,0546	30893	M+H-H ₂ O	2	132,0575	Marasin						C ₉ H ₈ O
115,0546	30895	M+H-H ₂ O	2	132,0575	3,4-octadien-5,7-diyn-1-ol						C ₉ H ₈ O
115,0546	157941	M+H-H ₂ O	2	132,0575	2-Indanone; 2H-Inden-2-one, 1,3-dihydro-						C ₉ H ₈ O
115,0546	0	M+H+HCOONa	0	0	No compounds found for experimental mass 115.05456 and adduct: M+H+HCOONa						
70,013	0	M+H	0	0	No compounds found for experimental mass 70.01299 and adduct: M+H						
70,013	0	M+Na	0	0	No compounds found for experimental mass 70.01299 and adduct: M+Na						
70,013	0	M+K	0	0	No compounds found for experimental mass 70.01299 and adduct: M+K						
70,013	0	M+H-H ₂ O	0	0	No compounds found for experimental mass 70.01299 and adduct: M+H-H ₂ O						
70,013	0	M+H+HCOONa	0	0	No compounds found for experimental mass 70.01299 and adduct: M+H+HCOONa						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

409,7553	0	M+H	0	0	No compounds found for experimental mass 409.75528 and adduct: M+H						
409,7553	0	M+Na	0	0	No compounds found for experimental mass 409.75528 and adduct: M+Na						
409,7553	0	M+K	0	0	No compounds found for experimental mass 409.75528 and adduct: M+K						
409,7553	0	M+H-H ₂ O	0	0	No compounds found for experimental mass 409.75528 and adduct: M+H-H ₂ O						
409,7553	0	M+H+HCOONa	0	0	No compounds found for experimental mass 409.75528 and adduct: M+H+HCOONa						
74,0604	113494	M+H	5	73,0528	Aminoacetone						C ₃ H ₇ NO
74,0604	47966	M+H	5	73,0528	N,N-Dimethylformamide				Drug	Food	C ₃ H ₇ NO
74,0604	54766	M+H	5	73,0528	3-Aminopropionaldehyde						C ₃ H ₇ NO
74,0604	111392	M+H	5	73,0528	2-(methylamino)acetaldehyde						C ₃ H ₇ NO
74,0604	159742	M+H	5	73,0528	Acetone oxime; Acetoxime						C ₃ H ₇ NO
74,0604	0	M+Na	0	0	No compounds found for experimental mass 74.0604 and adduct: M+Na						
74,0604	0	M+K	0	0	No compounds found for experimental mass 74.0604 and adduct: M+K						
74,0604	155998	M+H-H ₂ O	3	91,0633	3-Aminopropane-1,2-diol; (R)-3-Amino-1,2-propanediol						C ₃ H ₉ NO ₂
74,0604	0	M+H+HCOONa	0	0	No compounds found for experimental mass 74.0604 and adduct: M+H+HCOONa						

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

506,362	179871	M+H	3	505,3532	C-8 Ceramide-1-phosphate						C26H52NO6P
506,362	34339	M+H	3	505,3532	CerP(d18:1/8:0)						C26H52NO6P
506,362	163409	M+H	3	505,3532	PC(O-18:2(9Z,12Z)/0:0)[U]						C26H52NO6P
506,362	3029	M+H	3	505,3532	PC(P-18:1(9Z)/0:0)						C26H52NO6P
506,362	0	M+Na	0	0	No compounds found for experimental mass 506.36196 and adduct: M+Na						
506,362	0	M+K	0	0	No compounds found for experimental mass 506.36196 and adduct: M+K						
506,362	163438	M+H-H2O	3	523,3638	enantio-PAF C-16						C26H54NO7P
506,362	165818	M+H-H2O	3	523,3638	PAF C-16				Drug	Food	C26H54NO7P
506,362	2641	M+H-H2O	3	523,3638	PC(O-16:0/2:0)				Drug	Food	C26H54NO7P
506,362	163439	M+H-H2O	3	523,3638	PC(O-16:0/2:0)[S]						C26H54NO7P
506,362	163440	M+H-H2O	3	523,3638	PC(O-16:0/2:0)[U]						C26H54NO7P
506,362	163441	M+H-H2O	3	523,3638	PC(18:0/0:0)[S]						C26H54NO7P
506,362	163442	M+H-H2O	3	523,3638	PC(18:0/0:0)[U]						C26H54NO7P
506,362	163443	M+H-H2O	3	523,3638	PC(0:0/18:0)[S]						C26H54NO7P
506,362	163444	M+H-H2O	3	523,3638	PC(0:0/18:0)[U]						C26H54NO7P
506,362	2944	M+H-H2O	3	523,3638	PC(18:0/0:0)						C26H54NO7P
506,362	2970	M+H-H2O	3	523,3638	PC(0:0/18:0)						C26H54NO7P
506,362	164538	M+H-H2O	3	523,3638	PC(2:0/O-16:0)[U]						C26H54NO7P
506,362	4288	M+H-H2O	3	523,3638	PE(21:0/0:0)						C26H54NO7P
506,362	0	M+H+HCOONa	0	0	No compounds found for experimental mass 506.36196 and adduct: M+H+HCOONa						
603,5365	18283	M+H	3	602,5274	1-(14-methyl-pentadecanoyl)-2-(8-[3]-ladderane-octanyl)-sn-glycerol						C39H70O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

603,5365	185911	M+H	3	602,5274	FAHFA(22:3-(2-O-17:0)) 2-DTRAHHDA						C39H70O4
603,5365	185912	M+H	3	602,5274	FAHFA(22:3-(3-O-17:0)) 3-DTRAHHDA						C39H70O4
603,5365	185913	M+H	3	602,5274	FAHFA(22:3-(4-O-17:0)) 4-DTRAHHDA						C39H70O4
603,5365	185914	M+H	3	602,5274	FAHFA(22:3-(5-O-17:0)) 5-DTRAHHDA						C39H70O4
603,5365	185915	M+H	3	602,5274	FAHFA(22:3-(6-O-17:0)) 6-DTRAHHDA						C39H70O4
603,5365	185916	M+H	3	602,5274	FAHFA(22:3-(7-O-17:0)) 7-DTRAHHDA						C39H70O4
603,5365	185917	M+H	3	602,5274	FAHFA(22:3-(8-O-17:0)) 8-DTRAHHDA						C39H70O4
603,5365	185918	M+H	3	602,5274	FAHFA(22:3-(9-O-17:0)) 9-DTRAHHDA						C39H70O4
603,5365	185919	M+H	3	602,5274	FAHFA(22:3-(10-O-17:0)) 10-DTRAHHDA						C39H70O4
603,5365	185920	M+H	3	602,5274	FAHFA(22:3-(11-O-17:0)) 11-DTRAHHDA						C39H70O4
603,5365	185921	M+H	3	602,5274	FAHFA(22:3-(12-O-17:0)) 12-DTRAHHDA						C39H70O4
603,5365	185922	M+H	3	602,5274	FAHFA(22:3-(13-O-17:0)) 13-DTRAHHDA						C39H70O4
603,5365	185923	M+H	3	602,5274	FAHFA(22:3-(14-O-17:0)) 14-DTRAHHDA						C39H70O4
603,5365	185924	M+H	3	602,5274	FAHFA(22:3-(15-O-17:0)) 15-DTRAHHDA						C39H70O4
603,5365	185925	M+H	3	602,5274	FAHFA(22:3-(16-O-17:0)) 16-DTRAHHDA						C39H70O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

603,5365	186794	M+H	3	602,5274	FAHFA(24:3-(2-O-15:0)) 2-TTRAHPDA						C39H70O4
603,5365	186795	M+H	3	602,5274	FAHFA(24:3-(3-O-15:0)) 3-TTRAHPDA						C39H70O4
603,5365	186796	M+H	3	602,5274	FAHFA(24:3-(4-O-15:0)) 4-TTRAHPDA						C39H70O4
603,5365	186797	M+H	3	602,5274	FAHFA(24:3-(5-O-15:0)) 5-TTRAHPDA						C39H70O4
603,5365	186798	M+H	3	602,5274	FAHFA(24:3-(6-O-15:0)) 6-TTRAHPDA						C39H70O4
603,5365	186799	M+H	3	602,5274	FAHFA(24:3-(7-O-15:0)) 7-TTRAHPDA						C39H70O4
603,5365	186800	M+H	3	602,5274	FAHFA(24:3-(8-O-15:0)) 8-TTRAHPDA						C39H70O4
603,5365	186801	M+H	3	602,5274	FAHFA(24:3-(9-O-15:0)) 9-TTRAHPDA						C39H70O4
603,5365	186802	M+H	3	602,5274	FAHFA(24:3-(10-O-15:0)) 10-TTRAHPDA						C39H70O4
603,5365	186803	M+H	3	602,5274	FAHFA(24:3-(11-O-15:0)) 11-TTRAHPDA						C39H70O4
603,5365	186804	M+H	3	602,5274	FAHFA(24:3-(12-O-15:0)) 12-TTRAHPDA						C39H70O4
603,5365	186805	M+H	3	602,5274	FAHFA(24:3-(13-O-15:0)) 13-TTRAHPDA						C39H70O4
603,5365	186806	M+H	3	602,5274	FAHFA(24:3-(14-O-15:0)) 14-TTRAHPDA						C39H70O4
603,5365	181945	M+H	3	602,5274	1-(hexadecanoyl)-2-(8-[3]- ladderane-octanyl)-sn-glycerol						C39H70O4
603,5365	181951	M+H	3	602,5274	1-(14-methyl-pentadecanoyl)-2-(8- [1]-ladderane-octanyl)-sn-glycerol						C39H70O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

603,5365	181952	M+H	3	602,5274	1-(hexadecanoyl)-2-(8-[1]-ladderane-octanyl)-sn-glycerol						C39H70O4
603,5365	180185	M+H	3	602,5274	1-Hexadecyl-2-arachidonoyl-glycerol						C39H70O4
603,5365	18265	M+Na	7	580,5431	DG(O-16:0/18:1(9Z))						C37H72O4
603,5365	183898	M+Na	7	580,5431	FAHFA(17:0-(2-O-20:0)) 2-HDAHAA						C37H72O4
603,5365	183899	M+Na	7	580,5431	FAHFA(17:0-(3-O-20:0)) 3-HDAHAA						C37H72O4
603,5365	183900	M+Na	7	580,5431	FAHFA(17:0-(4-O-20:0)) 4-HDAHAA						C37H72O4
603,5365	183901	M+Na	7	580,5431	FAHFA(17:0-(5-O-20:0)) 5-HDAHAA						C37H72O4
603,5365	183902	M+Na	7	580,5431	FAHFA(17:0-(6-O-20:0)) 6-HDAHAA						C37H72O4
603,5365	183903	M+Na	7	580,5431	FAHFA(17:0-(7-O-20:0)) 7-HDAHAA						C37H72O4
603,5365	183904	M+Na	7	580,5431	FAHFA(17:0-(8-O-20:0)) 8-HDAHAA						C37H72O4
603,5365	183905	M+Na	7	580,5431	FAHFA(17:0-(9-O-20:0)) 9-HDAHAA						C37H72O4
603,5365	183906	M+Na	7	580,5431	FAHFA(17:0-(10-O-20:0)) 10-HDAHAA						C37H72O4
603,5365	183907	M+Na	7	580,5431	FAHFA(17:0-(11-O-20:0)) 11-HDAHAA						C37H72O4
603,5365	183908	M+Na	7	580,5431	FAHFA(17:0-(12-O-20:0)) 12-HDAHAA						C37H72O4
603,5365	183909	M+Na	7	580,5431	FAHFA(17:0-(13-O-20:0)) 13-HDAHAA						C37H72O4
603,5365	183910	M+Na	7	580,5431	FAHFA(17:0-(14-O-20:0)) 14-HDAHAA						C37H72O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

603,5365	183911	M+Na	7	580,5431	FAHFA(17:0-(15-O-20:0)) 15-HDAHAA						C37H72O4
603,5365	183912	M+Na	7	580,5431	FAHFA(17:0-(16-O-20:0)) 16-HDAHAA						C37H72O4
603,5365	183913	M+Na	7	580,5431	FAHFA(17:0-(17-O-20:0)) 17-HDAHAA						C37H72O4
603,5365	183914	M+Na	7	580,5431	FAHFA(17:0-(18-O-20:0)) 18-HDAHAA						C37H72O4
603,5365	183915	M+Na	7	580,5431	FAHFA(17:0-(19-O-20:0)) 19-HDAHAA						C37H72O4
603,5365	183432	M+Na	7	580,5431	FAHFA(15:0-(2-O-22:0)) 2-PDAHBA						C37H72O4
603,5365	183433	M+Na	7	580,5431	FAHFA(15:0-(3-O-22:0)) 3-PDAHBA						C37H72O4
603,5365	183434	M+Na	7	580,5431	FAHFA(15:0-(4-O-22:0)) 4-PDAHBA						C37H72O4
603,5365	183435	M+Na	7	580,5431	FAHFA(15:0-(5-O-22:0)) 5-PDAHBA						C37H72O4
603,5365	183436	M+Na	7	580,5431	FAHFA(15:0-(6-O-22:0)) 6-PDAHBA						C37H72O4
603,5365	183437	M+Na	7	580,5431	FAHFA(15:0-(7-O-22:0)) 7-PDAHBA						C37H72O4
603,5365	183438	M+Na	7	580,5431	FAHFA(15:0-(8-O-22:0)) 8-PDAHBA						C37H72O4
603,5365	183439	M+Na	7	580,5431	FAHFA(15:0-(9-O-22:0)) 9-PDAHBA						C37H72O4
603,5365	183440	M+Na	7	580,5431	FAHFA(15:0-(10-O-22:0)) 10-PDAHBA						C37H72O4
603,5365	183441	M+Na	7	580,5431	FAHFA(15:0-(11-O-22:0)) 11-PDAHBA						C37H72O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

603,5365	183442	M+Na	7	580,5431	FAHFA(15:0-(12-O-22:0)) 12-PDAHBA						C37H72O4
603,5365	183443	M+Na	7	580,5431	FAHFA(15:0-(13-O-22:0)) 13-PDAHBA						C37H72O4
603,5365	183444	M+Na	7	580,5431	FAHFA(15:0-(14-O-22:0)) 14-PDAHBA						C37H72O4
603,5365	183445	M+Na	7	580,5431	FAHFA(15:0-(15-O-22:0)) 15-PDAHBA						C37H72O4
603,5365	183446	M+Na	7	580,5431	FAHFA(15:0-(16-O-22:0)) 16-PDAHBA						C37H72O4
603,5365	185494	M+Na	7	580,5431	FAHFA(22:0-(2-O-15:0)) 2-BAHPDA						C37H72O4
603,5365	183447	M+Na	7	580,5431	FAHFA(15:0-(17-O-22:0)) 17-PDAHBA						C37H72O4
603,5365	185495	M+Na	7	580,5431	FAHFA(22:0-(3-O-15:0)) 3-BAHPDA						C37H72O4
603,5365	183448	M+Na	7	580,5431	FAHFA(15:0-(18-O-22:0)) 18-PDAHBA						C37H72O4
603,5365	185496	M+Na	7	580,5431	FAHFA(22:0-(4-O-15:0)) 4-BAHPDA						C37H72O4
603,5365	183449	M+Na	7	580,5431	FAHFA(15:0-(19-O-22:0)) 19-PDAHBA						C37H72O4
603,5365	185497	M+Na	7	580,5431	FAHFA(22:0-(5-O-15:0)) 5-BAHPDA						C37H72O4
603,5365	183450	M+Na	7	580,5431	FAHFA(15:0-(20-O-22:0)) 20-PDAHBA						C37H72O4
603,5365	185498	M+Na	7	580,5431	FAHFA(22:0-(6-O-15:0)) 6-BAHPDA						C37H72O4
603,5365	183451	M+Na	7	580,5431	FAHFA(15:0-(21-O-22:0)) 21-PDAHBA						C37H72O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

603,5365	185499	M+Na	7	580,5431	FAHFA(22:0-(7-O-15:0)) 7-BAHPDA						C37H72O4
603,5365	185500	M+Na	7	580,5431	FAHFA(22:0-(8-O-15:0)) 8-BAHPDA						C37H72O4
603,5365	185501	M+Na	7	580,5431	FAHFA(22:0-(9-O-15:0)) 9-BAHPDA						C37H72O4
603,5365	185502	M+Na	7	580,5431	FAHFA(22:0-(10-O-15:0)) 10-BAHPDA						C37H72O4
603,5365	185503	M+Na	7	580,5431	FAHFA(22:0-(11-O-15:0)) 11-BAHPDA						C37H72O4
603,5365	185504	M+Na	7	580,5431	FAHFA(22:0-(12-O-15:0)) 12-BAHPDA						C37H72O4
603,5365	185505	M+Na	7	580,5431	FAHFA(22:0-(13-O-15:0)) 13-BAHPDA						C37H72O4
603,5365	185506	M+Na	7	580,5431	FAHFA(22:0-(14-O-15:0)) 14-BAHPDA						C37H72O4
603,5365	184741	M+Na	7	580,5431	FAHFA(20:0-(2-O-17:0)) 2-AAHHDA						C37H72O4
603,5365	184742	M+Na	7	580,5431	FAHFA(20:0-(3-O-17:0)) 3-AAHHDA						C37H72O4
603,5365	184743	M+Na	7	580,5431	FAHFA(20:0-(4-O-17:0)) 4-AAHHDA						C37H72O4
603,5365	184744	M+Na	7	580,5431	FAHFA(20:0-(5-O-17:0)) 5-AAHHDA						C37H72O4
603,5365	184745	M+Na	7	580,5431	FAHFA(20:0-(6-O-17:0)) 6-AAHHDA						C37H72O4
603,5365	184746	M+Na	7	580,5431	FAHFA(20:0-(7-O-17:0)) 7-AAHHDA						C37H72O4
603,5365	184747	M+Na	7	580,5431	FAHFA(20:0-(8-O-17:0)) 8-AAHHDA						C37H72O4

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

603,5365	184748	M+Na	7	580,5431	FAHFA(20:0-(9-O-17:0)) 9-AAHHDA						C37H72O4
603,5365	184749	M+Na	7	580,5431	FAHFA(20:0-(10-O-17:0)) 10-AAHHDA						C37H72O4
603,5365	184750	M+Na	7	580,5431	FAHFA(20:0-(11-O-17:0)) 11-AAHHDA						C37H72O4
603,5365	184751	M+Na	7	580,5431	FAHFA(20:0-(12-O-17:0)) 12-AAHHDA						C37H72O4
603,5365	184752	M+Na	7	580,5431	FAHFA(20:0-(13-O-17:0)) 13-AAHHDA						C37H72O4
603,5365	184753	M+Na	7	580,5431	FAHFA(20:0-(14-O-17:0)) 14-AAHHDA						C37H72O4
603,5365	184754	M+Na	7	580,5431	FAHFA(20:0-(15-O-17:0)) 15-AAHHDA						C37H72O4
603,5365	184755	M+Na	7	580,5431	FAHFA(20:0-(16-O-17:0)) 16-AAHHDA						C37H72O4
603,5365	0	M+K	0	0	No compounds found for experimental mass 603.53651 and adduct: M+K						
603,5365	61832	M+H-H2O	3	620,538	DG(14:1n5/0:0/22:1n9)						C39H72O5
603,5365	86166	M+H-H2O	3	620,538	DG(16:0/0:0/20:2n6)						C39H72O5
603,5365	61100	M+H-H2O	3	620,538	DG(16:1n7/0:0/20:1n9)						C39H72O5
603,5365	62178	M+H-H2O	3	620,538	DG(18:1n7/0:0/18:1n9)						C39H72O5
603,5365	130550	M+H-H2O	3	620,538	DG(14:0/0:0/22:2n6)						C39H72O5
603,5365	86795	M+H-H2O	3	620,538	DG(18:1n7/0:0/18:1n7)						C39H72O5
603,5365	48662	M+H-H2O	3	620,538	DG(18:1n9/0:0/18:1n9)						C39H72O5
603,5365	64819	M+H-H2O	3	620,538	DG(22:1(13Z)/14:1(9Z)/0:0)						C39H72O5
603,5365	149328	M+H-H2O	3	620,538	DG(14:1(9Z)/22:1(13Z)/0:0)						C39H72O5
603,5365	110684	M+H-H2O	3	620,538	DG(18:2(9Z,12Z)/18:0/0:0)						C39H72O5

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

603,5365	17768	M+H-H ₂ O	3	620,538	DG(18:1(9Z)/18:1(9Z)/0:0)						C39H72O5
603,5365	17769	M+H-H ₂ O	3	620,538	DG(18:0/18:2(9Z,12Z)/0:0)[iso2]						C39H72O5
603,5365	17773	M+H-H ₂ O	3	620,538	DG(16:1(9Z)/20:1(11Z)/0:0)[iso2]						C39H72O5
603,5365	17774	M+H-H ₂ O	3	620,538	DG(16:0/20:2(11Z,14Z)/0:0)[iso2]						C39H72O5
603,5365	68484	M+H-H ₂ O	3	620,538	DG(20:1(11Z)/16:1(9Z)/0:0)						C39H72O5
603,5365	119185	M+H-H ₂ O	3	620,538	DG(20:2(11Z,14Z)/16:0/0:0)						C39H72O5
603,5365	65470	M+H-H ₂ O	3	620,538	DG(18:1(11Z)/18:1(11Z)/0:0)						C39H72O5
603,5365	18116	M+H-H ₂ O	3	620,538	DG(14:0/22:2(13Z,16Z)/0:0)[iso2]						C39H72O5
603,5365	61401	M+H-H ₂ O	3	620,538	DG(22:2(13Z,16Z)/14:0/0:0)						C39H72O5
603,5365	102646	M+H-H ₂ O	3	620,538	DG(18:1(11Z)/18:1(9Z)/0:0)						C39H72O5
603,5365	94463	M+H-H ₂ O	3	620,538	DG(18:1(9Z)/18:1(11Z)/0:0)						C39H72O5
603,5365	18244	M+H-H ₂ O	3	620,538	DG(19:1(9Z)/17:1(9Z)/0:0)[iso2]						C39H72O5
603,5365	17777	M+H-H ₂ O	3	620,538	DG(17:2(9Z,12Z)/19:0/0:0)[iso2]						C39H72O5
603,5365	18139	M+H-H ₂ O	3	620,538	DG(14:1(9Z)/22:1(11Z)/0:0)[iso2]						C39H72O5
603,5365	180216	M+H-H ₂ O	3	620,538	1,2-Dioleoyl-rac-glycerol						C39H72O5
603,5365	180652	M+H+HCOONa	8	534,5376	C36:1						C36H70O2
603,5365	31306	M+H+HCOONa	8	534,5376	Behenyl myristoleate						C36H70O2
603,5365	31307	M+H+HCOONa	8	534,5376	Myristoleyl behenate						C36H70O2
603,5365	31308	M+H+HCOONa	8	534,5376	Oleyl stearate						C36H70O2
603,5365	31309	M+H+HCOONa	8	534,5376	Palmitoleyl arachidate						C36H70O2
603,5365	31310	M+H+HCOONa	8	534,5376	Stearyl oleate						C36H70O2
603,5365	31156	M+H+HCOONa	8	534,5376	8E-octadecenyl heptadecanoate						C36H70O2
603,5365	31175	M+H+HCOONa	8	534,5376	Arachidyl palmitoleate						C36H70O2

Tabla A.1. Extracto de la tabla generada por el módulo *Tagger*, que muestra las 5 clasificaciones realizadas sobre las anotaciones de la tabla generada por Ceu Mass Mediator 3.0. La tabla completa se encuentra en la ruta /test/1_tagged_MS_experiment.xls del repositorio en GitHub del proyecto.

A.2. Tabla Generada por RENAME

Experimental mass	Adduct	mz Error (ppm)	Molecular Weight	Name	Peptide	Halogenated	Microbial	Drug	Food	Formula
336,8786	M+H-H ₂ O	7	353,8843	Bithionol		x				C ₁₂ H ₆ Cl ₄ O ₂ S
336,8786	M+H-H ₂ O	7	353,8843	Tetradifon		x				C ₁₂ H ₆ Cl ₄ O ₂ S
98,0577	M+Na	1	75,0684	Trimethylamine-oxide			MC			C ₃ H ₉ NO
98,0577	M+Na	1	75,0684	aminopropan-ol						C ₃ H ₉ NO
98,0577	M+Na	1	75,0684	(methylamino)ethan-ol						C ₃ H ₉ NO
189,0468	M+H	1	188,0393	Chlorobiphenyl		x				C ₁₂ H ₉ Cl
189,0468	M+Na	7	166,0565	Ethionamide				Drug		C ₈ H ₁₀ N ₂ S
189,0468	M+H-H ₂ O	7	206,0514	Indolylmethylthiohydroximate						C ₁₀ H ₁₀ N ₂ OS
189,0468	M+H-H ₂ O	7	206,0514	[(Furanylmethyl)thio]-methylpyrazine					Food	C ₁₀ H ₁₀ N ₂ OS
205,0244	M+H	9	204,0189	Honaucin		x				C ₈ H ₉ ClO ₄
205,0244	M+Na	9	182,0368	Acenaphthenequinone						C ₁₂ H ₆ O ₂
205,0244	M+H+HCOONa	5	136,0291	chloropentanoate		x				C ₅ H ₉ O ₂ Cl
117,055	M+H	3	116,0473	keto valerate						C ₅ H ₈ O ₃
117,055	M+H	3	116,0473	Acetoxyacetone					Food	C ₅ H ₈ O ₃
117,055	M+H	3	116,0473	Ethyl pyruvate						C ₅ H ₈ O ₃
117,055	M+H	3	116,0473	methyl-oxo-butanoate						C ₅ H ₈ O ₃
117,055	M+H	3	116,0473	hydroxy-methylbut-enoate						C ₅ H ₈ O ₃
117,055	M+H	3	116,0473	dimethyloxirane-carboxylate						C ₅ H ₈ O ₃
117,055	M+Na	1	94,0657	Methylpyridinium						C ₆ H ₈ N
117,055	M+H-H ₂ O	2	134,0579	Deoxyribose						C ₅ H ₁₀ O ₄

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

117,055	M+H-H ₂ O	2	134,0579	Deoxy-xylulose						C5H10O4
117,055	M+H-H ₂ O	2	134,0579	dihydroxy-valerate						C5H10O4
117,055	M+H-H ₂ O	2	134,0579	glycerol acetate						C5H10O4
117,055	M+H-H ₂ O	2	134,0579	hydroxy-(hydroxymethyl)-methylpropanoate						C5H10O4
117,055	M+H-H ₂ O	2	134,0579	Deoxy-arabinose						C5H10O4
117,055	M+H-H ₂ O	8	134,0592	Dihydropteridine						C6H6N4
95,0606	M+H	2	94,0531	Aminopyridine						C5H6N2
95,0606	M+H	2	94,0531	Methylpyrazine					Food	C5H6N2
162,0765	M+H	3	161,0688	Alanyl-lactate						C6H11NO4
162,0765	M+H	3	161,0688	Amino-dideoxy-scylo-inosose			MC			C6H11NO4
162,0765	M+H	3	161,0688	Methyl-glutamate						C6H11NO4
162,0765	M+H	3	161,0688	Iminobispropanoate					Food	C6H11NO4
162,0765	M+H	3	161,0688	Glutamate methylester						C6H11NO4
162,0765	M+H	3	161,0688	Acetylhomoserine						C6H11NO4
162,0765	M+H	3	161,0688	Acetylthreonine						C6H11NO4
162,0765	M+H	3	161,0688	Aminoadipate						C6H11NO4
162,0765	M+H	3	161,0688	hydroxybutyrylglycine						C6H11NO4
162,0765	M+H-H ₂ O	2	179,0794	Nojirimycin			MC			C6H13NO5
162,0765	M+H-H ₂ O	2	179,0794	Kanosamine			MC			C6H13NO5
162,0765	M+H-H ₂ O	2	179,0794	neo-Inosamine-2			MC			C6H13NO5
162,0765	M+H-H ₂ O	2	179,0794	Fructosamine						C6H13NO5
162,0765	M+H-H ₂ O	2	179,0794	Mannosamine						C6H13NO5
162,0765	M+H-H ₂ O	6	179,0807	Aminomethyl-carbaguanine						C7H9N5O
498,2905	M+Na	7	475,3046	MG132						C26H41N3O5

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

498,2905	M+Na	1	475,3006	Netilmicin				Drug		C21H41N5O7
498,2905	M+H-H2O	4	515,2917	Taurocholate						C26H45NO7S
498,2905	M+H-H2O	4	515,2917	Tauro-muricholate						C26H45NO7S
498,2905	M+H-H2O	4	515,2917	Tauroursobolate						C26H45NO7S
357,28	M+H	3	356,2715	tetracosahexaenoate						C24H36O2
357,28	M+H	3	356,2715	Chola-dien-oate						C24H36O2
357,28	M+H	3	356,2715	hydroxy-didehydro-trinorcalciol						C24H36O2
357,28	M+H	3	356,2715	THA				Drug		C24H36O2
357,28	M+H	3	356,2715	Docosahexaenoate ethyl ester						C24H36O2
357,28	M+H-H2O	3	374,2821	Hydroxy-chol-en-oate						C24H38O3
357,28	M+H-H2O	3	374,2821	(Heptadecenyl)-hydroxybenzoate					Food	C24H38O3
357,28	M+H-H2O	3	374,2821	(pentadec-en-yl)salicylate					Food	C24H38O3
357,28	M+H-H2O	3	374,2821	dihydroxy-trinorvitamin D3						C24H38O3
357,28	M+H-H2O	3	374,2821	Oxo-cholan-oate						C24H38O3
357,28	M+H-H2O	3	374,2821	hydroxy-tetracos-pentaenoate						C24H38O3
357,28	M+H-H2O	3	374,2821	[(Tetrahydro-pyran-yl)oxy]androst-en-ol						C24H38O3
357,28	M+H-H2O	3	374,2821	Dihydroxy-trimethyl-pregn-en-one						C24H38O3
357,28	M+H-H2O	6	374,2855	thio-Acetyl MAGE						C21H42O3S
410,3279	M+H	3	409,3192	PGF2a diethyl amide						C24H43NO4
410,3279	M+H	3	409,3192	Lumula						C24H43NO4
367,1815	M+Na	10	344,1889	TAN-67						C23H24N2O
367,1815	M+H-H2O	2	384,1838	JWH						C25H24N2O2
367,1815	M+H-H2O	2	384,1838	JWH naphthyl isomer						C25H24N2O2
367,1815	M+H+HCOONa	3	298,1879	Diphthamide						C13H24N5O3

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

387,3272	M+H	4	386,3185	hydroxy-deoxy-oxavitamin D3						C26H42O2
387,3272	M+H	4	386,3185	Shibate						C26H42O2
387,3272	M+H	4	386,3185	Phenyl-eicosanedione					Food	C26H42O2
387,3272	M+H	4	386,3185	enantioxyandrost-en-one						C26H42O2
387,3272	M+H-H2O	3	404,329	Dihomolithocholate						C26H44O3
387,3272	M+H-H2O	3	404,329	dihydroxy-norvitamin D3						C26H44O3
387,3272	M+H-H2O	3	404,329	dihydroxy-nor-secovitamin D3						C26H44O3
142,0268	M+H	2	141,0193	carboxypenta-dienoate						C6H5O4
142,0268	M+H	3	141,0191	Ethanolamine phosphate						C2H8NO4P
142,0268	M+H	3	141,0191	(Amino-hydroxyethyl)phosphonate						C2H8NO4P
142,0268	M+H	5	141,0188	Hydroxymethyl-furanoate						C6H5O4
142,0268	M+Na	4	119,0371	Diacylglycerol-PI-pool						C7H5NO
142,0268	M+K	3	103,0633	Propyl carbamate						C4H9NO2
142,0268	M+K	3	103,0633	amino-butanoate						C4H9NO2
142,0268	M+K	3	103,0633	HBA						C4H9NO2
142,0268	M+K	3	103,0633	Acetyethanolamine					Food	C4H9NO2
142,0268	M+K	3	103,0633	Dimethylglycine						C4H9NO2
142,0268	M+K	3	103,0633	Ethylglycine						C4H9NO2
142,0268	M+K	3	103,0633	Butyl nitrite						C4H9NO2
142,0268	M+K	3	103,0633	Methylalanine						C4H9NO2
311,2959	M+H	5	310,2872	Dimethyl-octenyl dimethyl-octadienoate						C20H38O2
311,2959	M+H	5	310,2872	Octadecenyl acetate						C20H38O2
311,2959	M+H	5	310,2872	eicosenoate						C20H38O2
311,2959	M+H	5	310,2872	Prostanoate skeleton						C20H38O2

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

311,2959	M+H	5	310,2872	octadecenoate						C20H38O2
311,2959	M+H	5	310,2872	Phytenoate						C20H38O2
311,2959	M+H	5	310,2872	Eicosanolide						C20H38O2
311,2959	M+H	5	310,2872	Icosanedione					Food	C20H38O2
311,2959	M+H	5	310,2872	Ethyl oleate						C20H38O2
311,2959	M+H-H2O	4	328,2977	methoxy-nonadecanoate						C20H40O3
311,2959	M+H-H2O	4	328,2977	hydroxy-eicosanoate						C20H40O3
311,2959	M+H-H2O	4	328,2977	Polyoxyethylene 40 monostearate						C20H40O3
311,2959	M+H-H2O	4	328,2977	hydroxyphytanate					Food	C20H40O3
115,0546	M+H-H2O	2	132,0575	Atropaldehyde						C9H8O
115,0546	M+H-H2O	2	132,0575	Indanone						C9H8O
115,0546	M+H-H2O	2	132,0575	Methylcumarone						C9H8O
115,0546	M+H-H2O	2	132,0575	Phenyl-propenal					Food	C9H8O
115,0546	M+H-H2O	2	132,0575	octadien-diyn-ol						C9H8O
74,0604	M+H	5	73,0528	Aminoacetone						C3H7NO
74,0604	M+H	5	73,0528	Dimethylformamide				Drug	Food	C3H7NO
74,0604	M+H	5	73,0528	Aminopropionaldehyde						C3H7NO
74,0604	M+H	5	73,0528	(methylamino)acetaldehyde						C3H7NO
74,0604	M+H	5	73,0528	Acetone oxime						C3H7NO
74,0604	M+H-H2O	3	91,0633	Aminopropane-diol						C3H9NO2
506,362	M+H	3	505,3532	CerP(26:1)						C26H52NO6P
506,362	M+H	3	505,3532	LPC(O-18:2)						C26H52NO6P
506,362	M+H	3	505,3532	LPC(P-18:1)						C26H52NO6P
506,362	M+H-H2O	3	523,3638	PAF C-16				Drug	Food	C26H54NO7P
506,362	M+H-H2O	3	523,3638	PC(O-18:0)				Drug	Food	C26H54NO7P

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

506,362	M+H-H ₂ O	3	523,3638	LPC(18:0)						C26H54NO7P
506,362	M+H-H ₂ O	3	523,3638	LPE(21:0)						C26H54NO7P
603,5365	M+H	3	602,5274	(methyl-pentadecanoyl)- (ladderane-octanyl)-glycerol						C39H70O4
603,5365	M+H	3	602,5274	FAHFA(39:3)						C39H70O4
603,5365	M+H	3	602,5274	(hexadecanoyl)-(ladderane- octanyl)-glycerol						C39H70O4
603,5365	M+H	3	602,5274	Hexadecyl-arachidonoyl-glycerol						C39H70O4
603,5365	M+Na	7	580,5431	DG(O-34:1)						C37H72O4
603,5365	M+Na	7	580,5431	FAHFA(37:0)						C37H72O4
603,5365	M+H-H ₂ O	3	620,538	DG(36:2)						C39H72O5
603,5365	M+H+HCOONa	8	534,5376	C36:1						C36H70O2
603,5365	M+H+HCOONa	8	534,5376	Myristoleyl behenate						C36H70O2
603,5365	M+H+HCOONa	8	534,5376	Stearyl oleate						C36H70O2
603,5365	M+H+HCOONa	8	534,5376	Arachidyl palmitoleate						C36H70O2
603,5365	M+H+HCOONa	8	534,5376	octadecenyl heptadecanoate						C36H70O2

Tabla A.2. Extracto de la tabla generada por el módulo *REname*. La tabla completa se encuentra en la ruta /test/2_REnamed_MS_experiment.xls del repositorio en GitHub del proyecto.

A.3. Tabla Generada por RowMerger

Experimental mass	Adduct	mz Error (ppm)	Molecular Weight	Name	Peptide	Halogenated	Microbial	Drug	Food	Formula
336,8786	M+H-H ₂ O	7	353,8843	Bithionol // Tetradifon		x				C ₁₂ H ₆ Cl ₄ O ₂ S
98,0577	M+Na	1	75,0684	Trimethylamine-oxide			MC			C ₃ H ₉ NO
98,0577	M+Na	1	75,0684	aminopropan-ol // (methylamino)ethan-ol						C ₃ H ₉ NO
189,0468	M+H	1	188,0393	Chlorobiphenyl		x				C ₁₂ H ₉ Cl
189,0468	M+Na	7	166,0565	Ethionamide				Drug		C ₈ H ₁₀ N ₂ S
189,0468	M+H-H ₂ O	7	206,0514	Indolylmethylthiohydroximate						C ₁₀ H ₁₀ N ₂ OS
189,0468	M+H-H ₂ O	7	206,0514	[(Furanylmethyl)thio]-methylpyrazine					Food	C ₁₀ H ₁₀ N ₂ OS
205,0244	M+H	9	204,0189	Honaucin		x				C ₈ H ₉ ClO ₄
205,0244	M+Na	9	182,0368	Acenaphthenequinone						C ₁₂ H ₆ O ₂
205,0244	M+H+HCOONa	5	136,0291	chloropentanoate		x				C ₅ H ₉ O ₂ Cl
117,055	M+H	3	116,0473	keto valerate // Ethyl pyruvate // methyl-oxo-butanoate // hydroxy-methylbut-enoate // dimethyloxirane-carboxylate						C ₅ H ₈ O ₃
117,055	M+H	3	116,0473	Acetoxyacetone					Food	C ₅ H ₈ O ₃
117,055	M+Na	1	94,0657	Methylpyridinium						C ₆ H ₈ N
117,055	M+H-H ₂ O	2	134,0579	Deoxyribose // Deoxy-xylulose // dihydroxy-valerate // glycerol acetate // hydroxy-(hydroxymethyl)-methylpropanoate // Deoxy-arabinose						C ₅ H ₁₀ O ₄
117,055	M+H-H ₂ O	8	134,0592	Dihydropteridine						C ₆ H ₆ N ₄

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

95,0606	M+H	2	94,0531	Aminopyridine						C5H6N2
95,0606	M+H	2	94,0531	Methylpyrazine					Food	C5H6N2
162,0765	M+H	3	161,0688	Alanyl-lactate // Methyl-glutamate // Glutamate methylester // Acetylhomoserine // Acetylthreonine // Aminoadipate // hydroxybutyrylglycine						C6H11NO4
162,0765	M+H	3	161,0688	Amino-dideoxy-scylo-inosose			MC			C6H11NO4
162,0765	M+H	3	161,0688	Iminobispropanoate					Food	C6H11NO4
162,0765	M+H-H2O	2	179,0794	Nojirimycin // Kanosamine // neo- Inosamine-2			MC			C6H13NO5
162,0765	M+H-H2O	2	179,0794	Fructosamine // Mannosamine						C6H13NO5
162,0765	M+H-H2O	6	179,0807	Aminomethyl-carbaguanine						C7H9N5O
498,2905	M+Na	7	475,3046	MG132						C26H41N3O5
498,2905	M+Na	1	475,3006	Netilmicin				Drug		C21H41N5O7
498,2905	M+H-H2O	4	515,2917	Taurocholate // Tauro-muricholate // Tauroursocholate						C26H45NO7S
357,28	M+H	3	356,2715	tetracosahexaenoate // Chola-dien- oate // hydroxy-didehydro- trinorcalciol // Docosahexaenoate ethyl ester						C24H36O2
357,28	M+H	3	356,2715	THA				Drug		C24H36O2
357,28	M+H-H2O	3	374,2821	Hydroxy-chol-en-oate // dihydroxy-trinorvitamin D3 // Oxo- cholan-oate // hydroxy-tetracosa- pentaenoate // [(Tetrahydro-pyran- yl)oxy]androst-en-ol // Dihydroxy- trimethyl-pregn-en-one						C24H38O3
357,28	M+H-H2O	3	374,2821	(Heptadecenyl)-hydroxybenzoate // (pentadec-en-yl)salicylate					Food	C24H38O3
357,28	M+H-H2O	6	374,2855	thio-Acetyl MAGE						C21H42O3S

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

410,3279	M+H	3	409,3192	PGF2a diethyl amide // Lumula						C24H43NO4
367,1815	M+Na	10	344,1889	TAN-67						C23H24N2O
367,1815	M+H-H2O	2	384,1838	JWH // JWH naphthyl isomer						C25H24N2O2
367,1815	M+H+HCOONa	3	298,1879	Diphthamide						C13H24N5O3
387,3272	M+H	4	386,3185	hydroxy-deoxy-oxavitamin D3 // Shibate // enanthoxyandrost-en-one						C26H42O2
387,3272	M+H	4	386,3185	Phenyl-eicosanedione					Food	C26H42O2
387,3272	M+H-H2O	3	404,329	Dihomolithocholate // dihydroxy-norvitamin D3 // dihydroxy-nor-secovitamin D3						C26H44O3
142,0268	M+H	2	141,0193	carboxypenta-dienoate						C6H5O4
142,0268	M+H	3	141,0191	Ethanolamine phosphate // (Amino-hydroxyethyl)phosphonate						C2H8NO4P
142,0268	M+H	5	141,0188	Hydroxymethyl-furanoate						C6H5O4
142,0268	M+Na	4	119,0371	Diacylglycerol-PI-pool						C7H5NO
142,0268	M+K	3	103,0633	Propyl carbamate // amino-butanoate // HBA // Dimethylglycine // Ethylglycine // Butyl nitrite // Methylalanine						C4H9NO2
142,0268	M+K	3	103,0633	Acetyethanolamine					Food	C4H9NO2
311,2959	M+H	5	310,2872	Dimethyl-octenyl dimethyl-octadienoate // Octadecenyl acetate // eicosenoate // Prostanate skeleton // octadecenoate // Phytanoate // Eicosanolide // Ethyl oleate						C20H38O2
311,2959	M+H	5	310,2872	Icosanedione					Food	C20H38O2
311,2959	M+H-H2O	4	328,2977	methoxy-nonadecanoate // hydroxy-eicosanoate // Polyoxyethylene 40 monostearate						C20H40O3

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

311,2959	M+H-H ₂ O	4	328,2977	hydroxyphytanate					Food	C ₂₀ H ₄₀ O ₃
115,0546	M+H-H ₂ O	2	132,0575	Atropaldehyde // Indanone // Methylcumarone // octadien-diyne-ol						C ₉ H ₈ O
115,0546	M+H-H ₂ O	2	132,0575	Phenyl-propenal					Food	C ₉ H ₈ O
74,0604	M+H	5	73,0528	Aminoacetone // Aminopropionaldehyde // (methylamino)acetaldehyde // Acetone oxime						C ₃ H ₇ NO
74,0604	M+H	5	73,0528	Dimethylformamide				Drug	Food	C ₃ H ₇ NO
74,0604	M+H-H ₂ O	3	91,0633	Aminopropane-diol						C ₃ H ₉ NO ₂
506,362	M+H	3	505,3532	CerP(26:1) // LPC(O-18:2) // LPC(P-18:1)						C ₂₆ H ₅₂ NO ₆ P
506,362	M+H-H ₂ O	3	523,3638	PAF C-16 // PC(O-18:0)				Drug	Food	C ₂₆ H ₅₄ NO ₇ P
506,362	M+H-H ₂ O	3	523,3638	LPC(18:0) // LPE(21:0)						C ₂₆ H ₅₄ NO ₇ P
603,5365	M+H	3	602,5274	(methyl-pentadecanoyl)-(ladderane-octanyl)-glycerol // FAHFA(39:3) // (hexadecanoyl)-(ladderane-octanyl)-glycerol // Hexadecyl-arachidonoyl-glycerol						C ₃₉ H ₇₀ O ₄
603,5365	M+Na	7	580,5431	DG(O-34:1) // FAHFA(37:0)						C ₃₇ H ₇₂ O ₄
603,5365	M+H-H ₂ O	3	620,538	DG(36:2)						C ₃₉ H ₇₂ O ₅
603,5365	M+H+HCOONa	8	534,5376	C36:1 // Myristoleyl behenate // Stearyl oleate // Arachidyl palmitoleate // octadecenyl heptadecanoate						C ₃₆ H ₇₀ O ₂

Tabla A.3. Extracto de la tabla generada por el módulo *RowMerger*. La tabla completa se encuentra en la ruta /test/3_rowMerged_MS_experiment.xls del repositorio en GitHub del proyecto

A.4. Tabla Generada por TableMerger

Feature	Experimental mass	RT [min]	Adduct	mz Error (ppm)	Molecular Weight	Name	Peptide	Halogenated	Microbial	Drug	Food
A01139	336,8786	0,338	M+H-H ₂ O	7	353,8843	Bithionol // Tetradifon		x			
A01283	98,0577	0,341	M+Na	1	75,0684	Trimethylamine-oxide			MC		
A01283	98,0577	0,341	M+Na	1	75,0684	aminopropan-ol // (methylamino)ethan-ol					
A01557	189,0468	0,349	M+H	1	188,0393	Chlorobiphenyl		x			
A01557	189,0468	0,349	M+Na	7	166,0565	Ethionamide				Drug	
A01557	189,0468	0,349	M+H-H ₂ O	7	206,0514	Indolylmethylthiohydroximate					
A01557	189,0468	0,349	M+H-H ₂ O	7	206,0514	[(Furanylmethyl)thio]-methylpyrazine					Food
A01654	205,0244	0,353	M+H	9	204,0189	Honaucin		x			
A01654	205,0244	0,353	M+Na	9	182,0368	Acenaphthenequinone					
A01654	205,0244	0,353	M+H+HCOONa	5	136,0291	chloropentanoate		x			
A01917	117,055	0,375	M+H	3	116,0473	keto valerate // Ethyl pyruvate // methyl-oxo-butanoate // hydroxy-methylbut-enoate // dimethyloxirane-carboxylate					
A01917	117,055	0,375	M+H	3	116,0473	Acetoxyacetone					Food
A01917	117,055	0,375	M+Na	1	94,0657	Methylpyridinium					
A01917	117,055	0,375	M+H-H ₂ O	2	134,0579	Deoxyribose // Deoxy-xylulose // dihydroxy-valerate // glycerol acetate // hydroxy-(hydroxymethyl)-methylpropanoate // Deoxy-arabinose					

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

A01917	117,055	0,375	M+H-H ₂ O	8	134,0592	Dihydropteridine					
A02180	95,0606	0,413	M+H	2	94,0531	Aminopyridine					
A02180	95,0606	0,413	M+H	2	94,0531	Methylpyrazine					Food
A02242	162,0765	0,424	M+H	3	161,0688	Alanyl-lactate // Methyl-glutamate // Glutamate methylester // Acetylhomoserine // Acetylthreonine // Amino adipate // hydroxybutyrylglycine					
A02242	162,0765	0,424	M+H	3	161,0688	Amino-dideoxy-scylo-inosose			MC		
A02242	162,0765	0,424	M+H	3	161,0688	Iminobispropanoate					Food
A02242	162,0765	0,424	M+H-H ₂ O	2	179,0794	Nojirimycin // Kanosamine // neo-Inosamine-2			MC		
A02242	162,0765	0,424	M+H-H ₂ O	2	179,0794	Fructosamine // Mannosamine					
A02242	162,0765	0,424	M+H-H ₂ O	6	179,0807	Aminomethyl-carbaguanine					
A05006	498,2905	6,644	M+Na	7	475,3046	MG132					
A05006	498,2905	6,644	M+Na	1	475,3006	Netilmicin				Drug	
A05006	498,2905	6,644	M+H-H ₂ O	4	515,2917	Taurocholate // Tauro-muricholate // Tauroursicholate					
A06251	357,28	8,885	M+H	3	356,2715	tetracosahexaenoate // Chola-dien-oate // hydroxy-didehydro-trinorcalciol // Docosahexaenoate ethyl ester					
A06251	357,28	8,885	M+H	3	356,2715	THA				Drug	
A06251	357,28	8,885	M+H-H ₂ O	3	374,2821	Hydroxy-chol-en-oate // dihydroxy-trinorvitamin D3 // Oxo-cholan-oate // hydroxy-tetracosapentaenoate // [(Tetrahydro-pyran-yl)oxy]androst-en-ol //					

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

						Dihydroxy-trimethyl-pregn-en-one					
A06251	357,28	8,885	M+H-H ₂ O	3	374,2821	(Heptadecenyl)-hydroxybenzoate // (pentadec-en-yl)salicylate					Food
A06251	357,28	8,885	M+H-H ₂ O	6	374,2855	thio-Acetyl MAGE					
A06301	410,3279	8,98	M+H	3	409,3192	PGF2a diethyl amide // Lumula					
A09682	367,1815	11,699	M+Na	10	344,1889	TAN-67					
A09682	367,1815	11,699	M+H-H ₂ O	2	384,1838	JWH // JWH naphthyl isomer					
A09682	367,1815	11,699	M+H+HCOONa	3	298,1879	Diphthamide					
B00517	387,3272	13,34	M+H	4	386,3185	hydroxy-deoxy-oxavitamin D3 // Shibate // enanthoxyandrost-en-one					
B00517	387,3272	13,34	M+H	4	386,3185	Phenyl-eicosanedione					Food
B00517	387,3272	13,34	M+H-H ₂ O	3	404,329	Dihomolithocholate // dihydroxy-norvitamin D3 // dihydroxy-nor-secovitamin D3					
B00652	142,0268	13,53	M+H	2	141,0193	carboxypenta-dienoate					
B00652	142,0268	13,53	M+H	3	141,0191	Ethanolamine phosphate // (Amino-hydroxyethyl)phosphonate					
B00652	142,0268	13,53	M+H	5	141,0188	Hydroxymethyl-furanoate					
B00652	142,0268	13,53	M+Na	4	119,0371	Diacylglycerol-PI-pool					
B00652	142,0268	13,53	M+K	3	103,0633	Propyl carbamate // amino-butanoate // HBA // Dimethylglycine // Ethylglycine // Butyl nitrite // Methylalanine					
B00652	142,0268	13,53	M+K	3	103,0633	Acetyethanolamine					Food
B01576	311,2959	14,617	M+H	5	310,2872	Dimethyl-octenyl dimethyl-octadienoate // Octadecenyl acetate // eicosenoate //					

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

						Prostanoate skeleton // octadecenoate // Phytienoate // Eicosanolide // Ethyl oleate					
B01576	311,2959	14,617	M+H	5	310,2872	Icosanedione					Food
B01576	311,2959	14,617	M+H-H ₂ O	4	328,2977	methoxy-nonadecanoate // hydroxy-eicosanoate // Polyoxyethylene 40 monostearate					
B01576	311,2959	14,617	M+H-H ₂ O	4	328,2977	hydroxyphytanate					Food
B02463	115,0546	15,983	M+H-H ₂ O	2	132,0575	Atropaldehyde // Indanone // Methylcumarone // octadien-diy-ol					
B02463	115,0546	15,983	M+H-H ₂ O	2	132,0575	Phenyl-propenal					Food
B04802	74,0604	18,088	M+H	5	73,0528	Aminoacetone // Aminopropionaldehyde // (methylamino)acetaldehyde // Acetone oxime					
B04802	74,0604	18,088	M+H	5	73,0528	Dimethylformamide				Drug	Food
B04802	74,0604	18,088	M+H-H ₂ O	3	91,0633	Aminopropane-diol					
B05982	506,362	19,368	M+H	3	505,3532	CerP(26:1) // LPC(O-18:2) // LPC(P-18:1)					
B05982	506,362	19,368	M+H-H ₂ O	3	523,3638	PAF C-16 // PC(O-18:0)				Drug	Food
B05982	506,362	19,368	M+H-H ₂ O	3	523,3638	LPC(18:0) // LPE(21:0)					
B08570	603,5365	25,173	M+H	3	602,5274	(methyl-pentadecanoyl)- (ladderane-octanyl)-glycerol // FAHFA(39:3) // (hexadecanoyl)- (ladderane-octanyl)-glycerol // Hexadecyl-arachidonoyl-glycerol					
B08570	603,5365	25,173	M+Na	7	580,5431	DG(O-34:1) // FAHFA(37:0)					
B08570	603,5365	25,173	M+H-H ₂ O	3	620,538	DG(36:2)					
B08570	603,5365	25,173	M+H+HCOONa	8	534,5376	C36:1 // Myristoleyl behenate // Stearyl oleate // Arachidyl					

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

						palmitoleate // octadecenyl heptadecanoate					
A01885	80,53221	0,371				UNK					
A02243	95,00101	0,424				UNK					
A05603	244,6389	7,622				UNK					
A08856	339,60535	11,172				UNK					
A10476	434,79218	12,116				UNK					
B02820	70,01299	16,01				UNK					
B03940	409,75528	17,37				UNK					

Tabla A.4. Extracto de la tabla generada por el módulo *TableMerger*. La tabla completa se encuentra en la ruta /test/4_tableMerged_MS_experiment.xls del repositorio en GitHub del proyecto

A.5. Revisión Manual de las Anotaciones

Feature	Nombre	Priorización
A01139	Bithionol // Tetradifon	Descartados por ser compuestos halogenados
A01283	Trimethylamine-oxide	Metabolito de origen microbiano que deberá ser revisado en posteriores análisis
	aminopropan-ol // (methylamino)ethan-ol	Ninguno de los metabolitos candidatos tiene un origen reportado en las bases de datos
A01557	Chlorobiphenyl	Descartado por ser compuesto halogenado
	Ethionamide	Descartado por ser fármaco no endógeno
	[(Furanylmethyl)thio]-methylpyrazine	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados.
	Indolylmethylthiohydroximate	Metabolito endógeno considerado más probable para esta <i>feature</i>
A01654	Chloropentanoate	Descartado por ser compuesto halogenado
	Honaucin	Descartado por ser compuesto halogenado
	Acenaphthenequinone	Descartado por ser compuesto empleado en la industria agroquímica y para tinciones
A01917	Acetoxyacetone	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados.
	Methylpyridinium	Compuesto que se forma durante el tueste del café

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

	Deoxyribose // Deoxy-xylulose // dihydroxy-valerate // glycerol acetate // hydroxy-(hydroxymethyl)-methylpropanoate // Deoxy-arabinose	Metabolitos endógenos
	keto valerate // Ethyl pyruvate // methyl-oxo-butanoate // hydroxy-methylbut-enoate // dimethyloxirane-carboxylate	Metabolitos endógenos
	Dihydropteridine	Metabolito endógeno
A02180	Aminopyridine	Descartado por tratarse de un fármaco contenido en la base de datos de DrugBank
	Methylpyrazine	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados
A02242	Nojirimycin // Kanosamine // neo-Inosamine-2	Metabolitos de origen microbiano que deberán ser revisados en posteriores análisis
	Amino-dideoxy-scylo-inosose	Metabolito de origen microbiano que deberá ser revisado en posteriores análisis
	Iminobispropanoate	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados
	Aminomethyl-carbaguanine	Metabolito endógeno
	Fructosamine // Mannosamine	Metabolitos endógenos
	Alanyl-lactate // Methyl-glutamate // Glutamate methylester // Acetylhomoserine // Acetylthreonine // Amino adipate // hydroxybutyrylglycine	<i>Alanyl-lactate</i> podría ser de origen microbiano. <i>Acetylthreonine</i> es detectado en placenta y heces, mientras que <i>hydroxybutyrylglycine</i> es detectado en orina. <i>Acetylhomoserine</i> y <i>Amino adipate</i> son metabolitos

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

		endógenos. <i>Methyl-glutamate</i> y <i>glutamate methylester</i> no tienen un origen reportado en las bases de datos
A05006	Netilmicin	Descartado por ser fármaco no endógeno
	MG132	Descartado por ser un inhibidor del proteosoma
	Taurocholate // Tauro-muricholate // Tauroursocolate	Metabolitos endógenos
A06251	THA	Descartado por ser fármaco no endógeno
	(Heptadecenyl)-hydroxybenzoate // (pentadec-en-yl)salicylate	(<i>pentadec-en-yl</i>) <i>salicylate</i> quedaría descartado por proceder de un fármaco no endógeno. (<i>Heptadecenyl</i>)- <i>hydroxybenzoate</i> procede de la dieta, por lo que se deberá comprobar si fue suministrado al organismo estudiado
	tetracosahexaenoate // Chola-dien-oate // hydroxy-didehydro-trinorcalciol // Docosahexaenoate ethyl ester	<i>Tetracosahexaenoate</i> y <i>chola-dien-oate</i> son metabolitos endógenos. <i>Hydroxy-didehydro-trinorcalciol</i> es un compuesto derivado de la vitamina D y <i>cocosahexaenoate ethyl ester</i> es un ácido graso esterificado. Estos dos últimos compuestos no tienen un origen reportado en las bases de datos
	Hydroxy-chol-en-oate // dihydroxy-trinorvitamin D3 // Oxo-cholan-oate // hydroxy-tetracosa-pentaenoate // [(Tetrahydro-pyran-yl)oxy]androst-en-ol // Dihydroxy-trimethyl-pregn-en-one	<i>Hydroxy-chol-en-oate</i> y <i>Oxo-cholan-oate</i> es un metabolito endógeno. <i>Dihydroxy-trinorvitamin D3</i> es un compuesto derivado de la vitamina D. <i>Hydroxy-tetracosa-pentaenoate</i> es un ácido graso insaturado. <i>Dihydroxy-trimethyl-pregn-en-one</i> es un compuesto derivado de hormonas. Estos tres últimos compuestos no tienen un origen reportado en las bases de datos
	thio-Acetyl MAGE	Descartado al ser un compuesto utilizado para medir actividad enzimática
A06301	PGF2a diethyl amide // Lumula	Compuestos sin un origen reportado en las bases de datos
A09682	JWH // JWH naphthyl isomer	Descartado por ser fármaco no endógeno
	(+)-TAN-67	Descartado por ser fármaco no endógeno

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

	Diphthamide	Metabolito más probable al formar parte del factor de elongación eucariota 2
B00517	Phenyl-eicosanedione	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados
	Dihomolithocholate // dihydroxy-norvitamin D3 // dihydroxy-nor-secovitamin D3	<i>dihydroxy-norvitamin D3</i> y <i>dihydroxy-nor-secovitamin D3</i> son compuestos derivados de la vitamina D. Ninguno de los compuestos candidatos tiene un origen reportado en las bases de datos
	hydroxy-deoxy-oxavitamin D3 // Shibate // enanthoxyandrost-en-one	Ninguno de los compuestos candidatos tiene un origen reportado en las bases de datos
B00652	carboxypenta-dienoate	No tiene un origen reportado en las bases de datos
	Ethanolamine phosphate // (Amino-hydroxyethyl)phosphonate	<i>Ethanolamine phosphate</i> es un metabolito endógeno. <i>(Amino-hydroxyethyl)phosphonate</i> no tiene un origen reportado en las bases de datos
	Propyl carbamate // amino-butanoate // HBA // Dimethylglycine // Ethylglycine // Butyl nitrite // Methylalanine	Todos han sido reportados como metabolitos endógenos, <i>excepto Propyl carbamate, HBA y Methylalanine</i>
	Acetylethanolamine	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados
	Diacylglycerol-PI-pool	Descartado al tratarse de un compuesto altamente tóxico
	Hydroxymethyl-furanoate	Metabolito endógeno

DESARROLLO DE HERRAMIENTAS COMPUTACIONALES PARA LA CLASIFICACIÓN DE METABOLITOS A PARTIR DE ANOTACIONES PUTATIVAS

B01576	methoxy-nonadecanoate // hydroxy-eicosanoate // Polyoxyethylene 40 monostearate	<i>Methoxy-nonadecanoate</i> ha sido detectado en <i>Helicobacter pylori</i> . <i>Hydroxy-eicosanoate</i> y <i>Polyoxyethylene 40 monostearate</i> son metabolitos endógenos
	hydroxyphytanate	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados
	Dimethyl-octenyl dimethyl-octadienoate // Octadecenyl acetate // eicosenoate // Prostanoate skeleton // octadecenoate // Phytienoate // Eicosanolide // Ethyl oleate	<i>Octadecenyl acetate</i> , <i>eicosenoate</i> , <i>octadecenoate</i> y <i>Ethyl oleate</i> son metabolitos endógenos. <i>Phytienoate</i> puede ser descartado al ser de origen vegetal. <i>Dimethyl-octenyl dimethyl-octadienoate</i> , <i>Prostanoate skeleton</i> y <i>Eicosanolide</i> no tienen un origen reportado en las bases de datos
	Icosanedione	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados
B02463	Atropaldehyde // Indanone // Methylcumarone // octadien-diyn-ol	<i>Indanone</i> y <i>Atropaldehyde</i> son metabolitos endógenos. <i>Methylcumarone</i> puede ser detectado en heces. <i>Octadien-diyn-ol</i> no tiene un origen reportado en las bases de datos
	Phenyl-propenal	Metabolito procedente de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados
B04802	Aminopropane-diol	No tienen un origen biológico reportado en las bases de datos
	Aminoacetone // Aminopropionaldehyde // (methylamino)acetaldehyde // Acetone oxime	Acetone oxime es un compuesto sintético. Los compuestos restantes son metabolitos endógenos
	Dimethylformamide	Metabolito procedente de fármaco o de la dieta. Se deberá comprobar si estaba presente en la dieta de los organismos estudiados
B05982	CerP(26:1) // LPC(O-18:2) // LPC(P-18:1)	Metabolitos endógenos

	PAF C-16 // PC(O-18:0)	Metabolitos procedentes de fármaco o de la dieta. Se deberá comprobar si estaban presentes en la dieta de los organismos estudiados
	LPC(18:0) // LPE(21:0)	LPC(18:0) es un metabolito endógeno. LPE(21:0) posiblemente sea de origen microbiano
B08570	(methyl-pentadecanoyl)-(ladderane-octanyl)-glycerol // FAHFA(39:3) // (hexadecanoyl)-(ladderane-octanyl)-glycerol // Hexadecyl-arachidonoyl-glycerol	No tienen un origen reportado en las bases de datos
	DG(36:2)	Metabolito endógeno
	DG(O-34:1) // FAHFA(37:0)	DG(O-34:1) es un metabolito endógeno. FAHFA(37:0) no tiene un origen reportado en las bases de datos
	C36:1 // Myristoleyl behenate // Stearyl oleate // octadecenyl heptadecanoate // Arachidyl palmitoleate	No tienen un origen reportado en las bases de datos

Tabla A.5. Priorización y descarte de las posibles identificaciones asignadas a las distintas *features*.

Apéndice B

Manual de Instalación

Los diferentes lanzamientos de la aplicación se pueden descargar desde [GitHub](https://github.com/CNIC-Proteomics/TurboPutative/releases) (<https://github.com/CNIC-Proteomics/TurboPutative/releases>). Así, la aplicación de escritorio está disponible para Windows (TurboPutative-x.x.x-pyenv-win32-x64.zip) y para Linux (TurboPutative-x.x.x-pyenv-linux-x64.zip). La versión “*pyenv*” tiene los programas de Python sin compilar, por lo que será necesario tener instalado Python 3.6 o una versión superior para poder utilizarla (<https://www.python.org/downloads/>). No obstante, también está disponible una versión (“*pycompiled*”) que, pese a ocupar más espacio, contiene los módulos de Python compilados, por lo que no requiere tener Python instalado en el equipo (TurboPutative-x.x.x-pycompiled-win32-x64.zip y TurboPutative-x.x.x-pycompiled-linux-x64.zip).

Para empezar a utilizar la versión de Windows es necesario ejecutar el *script* TurboPutative.bat. La primera vez que se ejecute la versión no compilada se iniciará la creación del entorno virtual de Python con las librerías requeridas (Numpy, Pandas, Cython, xlrd, xlwt y openpyxl). En caso de no encontrar Python en el PATH del sistema (o si este no es Python 3.6 o una versión superior), se pedirá al usuario que introduzca la dirección completa hacia un ejecutor de Python válido. El entorno virtual se creará en la carpeta *env* del directorio raíz de la aplicación. A continuación, se podrá iniciar la aplicación de escritorio ejecutando TurboPutative.bat.

Para utilizar TurboPutative en Linux es necesario ejecutar desde una terminal el *script* TurboPutative.sh (o utilizar el intérprete de Bash: `bash TurboPutative.sh`). La primera vez que se ejecute la versión no compilada se creará el entorno virtual de Python, de un modo similar al descrito para la versión de Windows. Tras la creación del entorno virtual se podrá usar la aplicación de escritorio ejecutando TurboPutative.sh.

Finalmente, es posible descargar el código fuente del servidor en TurboPutative-x.x.x-webServer.zip. Para iniciar el servidor web como *localhost* es necesario tener instalado en el equipo Node.js y su gestor de paquetes npm (<https://nodejs.org/es/>). Asimismo, para poder ejecutar flujos de trabajo es preciso que el servidor esté alojado en Linux. Tras descomprimir el fichero .zip descargado, se debe ejecutar con la terminal y desde la carpeta raíz el comando `npm install`, para descargar los paquetes y dependencias

requeridos. Para iniciar el servidor se debe ejecutar con la terminal y también desde la carpeta raíz del proyecto el comando `npm start`. A continuación, el servidor local escuchará las conexiones en el puerto por defecto 8080, y se podrán enviar solicitudes introduciendo en el navegador <http://localhost:8080/>.

Apéndice C

Organización del Código Fuente

El código fuente del proyecto es accesible a través de [GitHub](https://github.com/CNIC-Proteomics/TurboPutative) (<https://github.com/CNIC-Proteomics/TurboPutative>). Su contenido se organiza en tres carpetas principales contenidas en el directorio raíz: *desktopApp*, *webApp* y *tools*. La carpeta *desktopApp* contiene los ficheros y el código correspondiente a la aplicación de escritorio en Windows y en Linux. La carpeta *webApp* contiene los ficheros y el código de la aplicación web. Finalmente, la carpeta *tools* contiene algunas herramientas y otros ficheros utilizados para el desarrollo de TurboPutative. A continuación, explicamos en detalle el contenido de estos directorios.

Carpeta tools

La carpeta *tools* contiene un fichero `errorCode.json` que asigna una determinada información a algunos de los posibles códigos de error generados por el programa, a saber: el módulo que generó el error, descripción del error y su código asociado. Esta información será utilizada por la aplicación de escritorio y el servidor web para informar debidamente al usuario en caso de que se produjera un error en el flujo de trabajo.

La carpeta *tools* contiene cuatro directorios correspondientes a cada uno de los módulos que componen TurboPutative. En estos directorios podemos encontrar el programa en Python del módulo y un fichero en formato INI desde el cual el programa tomará los parámetros para la ejecución.

En la carpeta de *REname* es posible encontrar el fichero con las expresiones regulares (`regex.ini`) utilizadas por el módulo para procesar el nombre de los compuestos.

En la carpeta de *Tagger* hay una subcarpeta denominada *dbs*, que contiene las listas de nutrientes, fármacos y compuestos microbianos (`food_list.tsv`, `drug_list.tsv`, `microbial_list.tsv`) utilizadas por *Tagger* para realizar la clasificación de los compuestos. Asimismo, esta carpeta contiene un conjunto de *scripts* en Bash y Python que permiten generar estas listas a partir del fichero XML con todos los metabolitos de HMDB (<https://hmdb.ca/downloads>). Además, se obtienen todos los sinónimos de los

compuestos extraídos mediante acceso programático a PubChem (getAllSynonyms.py, dentro de la carpeta scripts).

Carpeta desktopApp

La carpeta *desktopApp* contiene los ficheros package.json y package-lock.json, que permiten instalar los paquetes y las dependencias de Node.js requeridos para el desarrollo y la ejecución de la aplicación de escritorio. Para más información, consultar las instalaciones mediante npm (<https://docs.npmjs.com/cli/v6/commands/npm-install>).

Los ficheros TurboPutative.bat y TurboPutative.sh permiten iniciar la ejecución de la aplicación en Windows y Linux, respectivamente. La carpeta *install* contiene los ficheros install_win32.bat e install_linux.sh, que permiten crear un entorno virtual de Python desde el que se ejecutarán los módulos .py de la aplicación. Para más información, consultar “[12. Virtual Environments and Packages — Python 3.6.12 documentation](#)”.

La carpeta *app* contiene la aplicación que se ejecutará con Electron empleando los binarios precompilados (ver *Manual distribution* en “[Application Distribution | Electron \(electronjs.org\)](#)”). En el interior de *app* se encuentra el fichero index.js, correspondiente al proceso principal y desde el que se abrirá la interfaz gráfica de usuario. Además, en el interior de *app* podemos encontrar cuatro directorios:

- assets: La carpeta *assets* contiene las hojas de estilo en cascada (CSS), imágenes y otros ficheros utilizados para el desarrollo de la interfaz de la aplicación.
- sections: La carpeta *sections* contiene los ficheros HTML (además de los ficheros CSS y JS asociados) de las dos principales secciones de la aplicación: *Help* y *Execute*. La sección *Help* permite al usuario consultar la documentación del programa, mientras que desde la sección *Execute* es posible diseñar un flujo de trabajo personalizado.
- jobs: La carpeta *jobs* contiene los directorios con los resultados de los flujos de trabajo ejecutados por el usuario.
- src: La carpeta *src* contiene el código y los ficheros necesarios para la ejecución del flujo de trabajo. En las siguientes líneas describimos con más detalle el contenido de este directorio.

En la carpeta *pyModule* de *src* se encuentran los programas en Python correspondientes a cada uno de los módulos. Las secuencias de comandos *integrator.bat* (versión de Windows) e *integrator.sh* (versión de Linux) integran los cuatro módulos de Python. Así, el fichero *index.js* que mencionamos anteriormente ejecutará el *script* *integrator*, que a su vez lanzará cada uno de los programas en Python en el orden apropiado y siguiendo las instrucciones definidas por el usuario. La carpeta *Data* de *src* contiene los ficheros utilizados por cada uno de los módulos, mientras que la carpeta *config* contiene ejemplos de ficheros de configuración en formato INI desde los cuales los programas podrán leer los parámetros. Sin embargo, estos no serán los ficheros de configuración utilizados por los módulos, ya que cada vez que el usuario inicie un flujo de trabajo se crearán unos ficheros INI con los nuevos parámetros en la carpeta *jobs* correspondiente. Estos ficheros de configuración serán los leídos por los programas.

Carpeta webApp

La carpeta *webApp* contiene los ficheros *package.json* y *package-lock.json*, necesarios para instalar los paquetes y las dependencias de Node.js requeridas para el desarrollo y la activación del servidor. Asimismo, en el interior de *webApp* podemos encontrar el directorio *src*, que contiene los distintos ficheros y el código utilizado por la aplicación web. En las líneas siguientes explicamos con detalle el contenido de *src*:

- *src/index.js*: El fichero *index.js* es el archivo principal de la aplicación web, desde el cual se inicia el servidor y se escuchan las conexiones.
- *src/routes/*: En la carpeta *routes* se encuentran los ficheros *index.js* y *execute.js*, en los que se especifica la gestión de los direccionamientos. Es en ellos donde se establece cómo debe responder la aplicación a una solicitud de cliente con un punto de acceso (URI) y un método de solicitud HTTP determinados. Así, la solicitud con punto de acceso “/turboputative.html” y método POST conduce a la ejecución del flujo de trabajo de acuerdo con los parámetros definidos por el usuario.
- *src/TurboPutative/*: La carpeta *TurboPutative* de *src* contiene el código necesario para la ejecución del flujo de trabajo. Así, *pyModules* contiene

los programas correspondientes a cada uno de los módulos. Para iniciar la ejecución, `route.js` ejecuta `integrator.sh`, que a su vez lanzará los programas en el orden determinado por el usuario. La carpeta *Data* contiene los ficheros utilizados por los diferentes módulos, mientras que en la carpeta *config* podemos encontrar ejemplos de los ficheros de configuración en formato INI, que utilizan los programas para leer los parámetros.

- `src/public/`: El directorio *public* contiene los ficheros estáticos del servidor. En su interior se encuentran los documentos HTML correspondientes a las diferentes secciones de la web. La carpeta *assets* contiene hojas de estilo en cascada (CSS), código de JavaScript y otros ficheros utilizados por los diferentes documentos HTML. Asimismo, la carpeta *images* contiene las imágenes utilizadas por la web.
- `src/views`: La carpeta *views* contiene los documentos HTML correspondientes a las diferentes secciones del sitio web.
- `src/partials/`: La carpeta *partials* contiene los documentos HTML que sirven de plantilla para generar las diferentes páginas del servidor.
- `src/lib/`: La carpeta *lib* contiene diferentes funciones de JavaScript utilizadas por el servidor.